

Capítulo III.

Puntos básicos de la evaluación: Metodología

Tema A.

¿Cuándo podemos decir que los resultados de un estudio son válidos y aplicables ?

Tema B.

¿Que decisión tomar si el nuevo medicamento es más eficaz o seguro que el de referencia, pero también es más caro ?.

Tema C.

¿Cuándo podemos decir que el nuevo medicamento es un equivalente terapéutico?

Tema D.

¿ Como definir condiciones de uso y lugar en terapéutica?

Capítulo III

TEMA A:

¿CUÁNDO PODEMOS DECIR QUE LOS RESULTADOS DE UN ESTUDIO SON VÁLIDOS Y APLICABLES ?

1-La calidad de los ensayos clínico publicados: Las evidencias bajo sospecha

2-Aprender a leer entre líneas.

2.1 .-Validez interna

A-Aspectos básicos

B-Aspectos secundarios

2.2.-Eficacia y validez externa o aplicabilidad: hacia la efectividad

2.3 Relevancia de los resultados

A-Concepto y cálculo RAR, NNT, RRR, RR, OR

B-Precisión y IC

3.-Decidir en la práctica

4. Anexo

-Lista Consort

1-LA CALIDAD DE LOS ENSAYOS CLÍNICOS PUBLICADOS: LAS EVIDENCIAS BAJO SOSPECHA

La validez interna y la aplicabilidad de los resultados de los estudios es el punto crucial para la selección del fármaco.

La lectura superficial de muchos ensayos clínicos pivotaes nos aporta evidencias sustanciales de las ventajas de un nuevo fármaco. Pero un análisis en profundidad nos pone de manifiesto aspectos críticos que matizan y a veces invalidan estas ventajas. El hecho de que la autorización para comercializar se base en un único o en unos pocos ensayos clínicos fase III, hace que el análisis pormenorizado de dicho(s) ensayo(s) sea especialmente decisivo para la toma de decisiones.

La presencia y detección de sesgos en los ensayos clínicos publicados es un tema cada vez más presente en la literatura médica.

Como ejemplos de la necesidad de esta lectura crítica y rigurosa, en los últimos meses, tenemos varios ejemplos de "evidencias" que han sufrido profundas revisiones por parte de la comunidad científica, por motivos diferentes. Los ejemplos recientes (ver tabla 1) :

- Estudio CLASS, celecoxib en osteoartritis y artritis reumatoide
- Estudio PROWESS, Drotrecogina alfa activada en sepsis
- Estudio WHI, terapia hormonal sustitutiva en mujer postmenopausica
- Rasburicasa en hiperuricemia postquimioterapia

La comunidad científica está muy sensibilizada por este tema, y como ejemplo de algunos de los esfuerzos y propuestas encaminadas a mejorar la calidad de los ensayos clínicos y de su publicación está **la lista-guía CONSORT**, aceptada por los editores de las revistas médicas de mayor impacto (Annals Intern Med, JAMA, NEJM, Lancet,...). Ha sido recientemente actualizada y constituye una herramienta excelente como guía para la revisión crítica de un artículo. (Altman DG; Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, Gotzsche PC, Lang T: Tehe revised CONSORT statment for reporting randomized trials: explanation and elaboration. Ann Intern Med 2001; 1134:663-694). Se puede acceder al protocolo de evaluación original, información ampliada y ejemplos en internet, acceso libre en <http://www.consort-statement.org/revisedstatement.htm>. Ver en anexo check-list modelo de la lista CONSORT.

Financiación, autoría y responsabilidad. Igualmente es motivo de preocupación el tema de la autoría de las publicaciones y de la relación de los autores con las compañías promotoras del ensayo. También ha generado una normativa y unos requisitos a tener en cuenta para la presentación de manuscritos para publicación en revistas biomédicas: Frank Davidoff, Catherine D. DeAngelis, Jeffrey M. Drazen, et al. Financiación, autoría y responsabilidad. Versión es español de la Rev Esp Cardiol. Acceso libre en Rev Esp Cardiol 2001; 54: 1247 - 1250. <http://www.secardiologia.es/>

Tabla 1
Ejemplos recientes de Ensayos Clínico y “evidencias” que se han puesto en duda

Publicación datos parciales

Como ejemplo de datos publicados de forma parcial y que pueden modificar la estimación de una relación riesgo-beneficio, está el estudio CLASS (Silverstein Fe et al . JAMA 2000; 284: 1246-5512) que comparaba celecoxib, diclofenaco e ibuprofeno en la indicación de osteoartritis y artritis reumatoide. Concluía que celecoxib se asociaba a una menor incidencia de úlceras sintomáticas y complicaciones ulcerosas que los AINEs clásicos. En una editorial reciente se analizan los datos del estudio completo y explica que el estudio original tenía significativos defectos de diseño: combinó el hallazgo de dos estudios distintos e incluyó los resultados de sólo los seis primeros meses del ensayo. Cuando se analizaron los datos de todo el periodo de 12 meses, celecoxib no produjo menos complicaciones ulcerosas que los dos AINEs clásicos. (Jüni P et al BMJ 2002; 324:1287-8)3.

Modificación del protocolo del ensayo clínico

Otro ejemplo es el del estudio PROWESS sobre Proteína C activada en la sepsis . El estudio sufrió una modificación en los criterios de inclusión durante el ensayo. Con 720 pacientes ya incluidos se decidió excluir a aquellos que presentaban riesgos mayores de muerte no ligados a sepsis y se procedió a cambiar la línea celular productora de la Proteína C activada, así como incluir nuevos criterios de análisis. La diferencia en mortalidad no era estadísticamente significativa en el subgrupo de pacientes respondiendo a los criterios iniciales (28% vs 30%) mientras que sí era significativa en los pacientes incluidos tras el cambio de protocolo (22% vs 31%) (Siegel JP NEJM Sounding Board; 2002; 347: 1030-4). Un comité de expertos (Anti-infective Drugs advisory Committee) de la FDA de los cuales la mitad de los miembros votaron a favor y la otra mitad en contra de su aprobación, procedió a reevaluar el ensayo clínico en función de la escala de valoración APACHE II obteniéndose información adicional. El análisis post-hoc han evidenciado subgrupos de enfermos que podrían obtener un beneficio de la Drotrecogina alfa y otros que estarían expuestos a riesgos de sangrado grave.

Nuevas evidencias que cambian las cosas

El tercer ejemplo es el de la Terapia Hormonal Sustitutiva (THS) con nuevos ensayos clínicos y que demuestran unos riesgos que no se conocían. Recientemente se han publicado los resultados del ensayo clínico WHI (Writing Group for the Women’s Health Initiative Investigators JAMA 2002; 288: 321-333), realizado en 16.608 mujeres con el objetivo de determinar los riesgos y beneficios a largo plazo de la THS en la mujer postmenopausica. Sus resultados muestran mayores riesgos de cáncer de mama y de tipo cardiovascular, que los que se conocían hasta el momento. El riesgo que se había estimado previamente a partir de estudios observacionales, mostraba ventajas de la THS en la disminución de riesgo cardiovascular. Los resultados de este ensayo clínico han modificado la relación beneficio/riesgo de la THS en prevención primaria (Fletcher WS et al . JAMA 2002; 298, 3: 366-), generando las advertencias oportunas por parte de las agencias reguladoras para limitar sus uso y advertir de los riesgos cardiovasculares a corto plazo y de neoplasia de mama a largo plazo.

Ensayos pequeños que “justifican” un registro.

La Rasburicasa fue autorizada por al EMEA en base 5 estudios de los cuales solo uno es un ensayo clínico de fase III. (Informe EMEA 2001). Dicho estudio (Goldman SC et al Blood 2001; 97:2998-3003) es un ensayo no ciego con un tamaño de muestra muy pequeño (52 pacientes pediátricos) que compara Alopurinol y Rasburicasa. El resultado se expresa en una variable intermedia (area bajo la curva de ácido úrico durante 96 horas). El estudio tiene un tamaño de muestra demasiado pequeño para encontrar diferencias de incidencia en fallo renal o requisitos de soporte renal. El hecho es que con estas evidencias el producto recibió autorización para ser comercializado

2-APRENDER A LEER ENTRE LÍNEAS.

La validez y aplicabilidad de los resultados de un ensayo debe revisarse haciendo teniendo en cuenta los conceptos y aspectos metodológico siguientes:

-La **validez interna** se plantea si los resultados obtenidos proporcionan una respuesta correcta a la pregunta planteada en el objetivo del estudio, con ausencia de sesgos en los proceso de diseño, ejecución y análisis.

-La **validez externa** de un estudio nos informa sobre la aplicabilidad de las conclusiones del mismo a otros pacientes distintos de los que participaron. Para determinar la validez externa de un estudio 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, etc . Se trata de ver si estos aspectos son los adecuados al problema de salud de interés.

2.1 VALIDEZ INTERNA DE UN ENSAYO CLÍNICO

Podemos decir que un ensayo clínico es válido si permite asegurar que el efecto observado es debido a la diferente intervención que han recibido el grupo intervención y el grupo control y no a otros factores que puedan haber conducido a una falsa conclusión .

A) ASPECTOS BÁSICOS

Se consideran tres aspectos básicos para que un ensayo clínico tenga un buen nivel de validez:

- 1-Asignación aleatoria
- 2-Seguimiento exhaustivo de todos los sujetos
- 3-Análisis por intención de tratar.

-**Asignación aleatoria.** La característica metodológica clave de un ensayo clínico es la asignación aleatoria de los sujetos a los grupos de estudio: grupo control y grupo intervención. Al intervenir solamente el azar en la asignación, se tiende a asegurar que se produce una distribución equilibrada de todas las variables, tanto conocidas como desconocidas. En estudios con un tamaño de muestra reducido, pueden emplearse las técnicas de asignación estratificada o por bloques que permiten aumentar la probabilidad de formar grupos equilibrados.

-**Seguimiento exhaustivo de todos los sujetos.** Se trata de responder a dos cuestiones: ¿se ha tenido en cuenta a todos los pacientes en los resultados ?, y ¿se ha realizado un seguimiento completo de los pacientes ?. 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 sujetos que finalizan el estudio

pueden formar grupos que no sean similares, de forma que la comparación de las respuestas observadas no sea imparcial.

De todas formas en la práctica siempre existe un número de pacientes que no pueden ser seguidos. Para poder evaluar este aspecto, es importante que en la publicación del ensayo clínico se proporcione **información sobre los abandonos (dropouts)** o pacientes que interrumpen el seguimiento y sobre las **retiradas (withdrawals)** o sujetos que deben abandonar por la aparición de reacciones adversas u otros sucesos que obligan a su exclusión. Si sus características son diferentes de la de los sujetos 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.

Actualmente las normas de publicación en revistas científicas siguen las recomendaciones del grupo CONSORT, y suelen presentar un diagrama de flujo donde constan las pérdidas de pacientes de cada fase del ensayo 1-Reclutamiento, 2-Asignación, 3-Seguimiento, 4-Análisis

-Análisis por la intención de tratar. Los investigadores pueden estar tentados de excluir determinados sujetos del análisis porque no han recibido el tratamiento previsto, han tenido un mal cumplimiento o no han seguido la pauta de visitas u otros aspectos del protocolo del estudio. Sin embargo, dado que los motivos por los que se han producido estos hechos pueden estar relacionados con el pronóstico de los sujetos, su exclusión podría introducir un sesgo de comparación. Para evitarlo, debe utilizarse la estrategia del **análisis por intención de tratar**, según la cual los sujetos se incluyen en el grupo al que han sido asignados independientemente de que si han recibido o no correctamente la intervención, seguimiento y evaluación

Conceptos:

-Análisis por intención de tratar. "Intention-to-treat analyses" (ITT). Según la definición más ampliamente aceptada es el análisis que incluye "a todos los pacientes en los grupos a los que fueron aleatoriamente asignados, independientemente de que cumplieran o no los criterios de inclusión, independientemente del tratamiento que realmente recibieran e independientemente de que abandonaran el tratamiento o se desviaran del protocolo". Algunos ensayos utilizan un concepto un poco más restrictivo, ejemplo: "Análisis que se realiza sobre los pacientes reclutados y aleatorizados que han recibido al menos una dosis del medicamento correcto según aleatorización".

-Análisis por intención de tratar modificado. "Modified Intention-to-treat analyses" (MITT): Muchos ensayos definen en la metodología que se realiza MITT. Por ejemplo el estudio (*Herbrecht R et al N Eng J Med 2002;347: 408-15*) se asigna tratamiento antifúngico (Voriconazol o Anfotericina B) cuando hay sospecha de infección por hongos. El MITT se analizan los pacientes que cumplen los criterios del análisis "por intención de tratar" y además presentan diagnóstico confirmado definitivo o probable de infección fúngica realizado por un grupo

independiente. Este tipo de análisis debe estar previsto y descrito en el método. Los análisis post-hoc tienen un valor muy limitado.

-Análisis "per protocol" (PP). Estudian sólo los pacientes de ambos grupos que han seguido totalmente el protocolo. (Ejemplo, se tienen sólo en cuenta los pacientes sin violaciones del protocolo, que cumplen todos los criterios de inclusión y exclusión, que han recibido el tratamiento correspondiente, adecuada evaluación al final del tratamiento, cumplimiento del tratamiento correcto). En ocasiones se presentan análisis de tipo PP adaptados como por ejemplo los llamados "análisis de pacientes con criterios de evaluación" definidos en la sección de método del ensayo.

En general los resultados ITT serán más similares a los que podemos esperar cuando se aplica el fármaco en clínica que los análisis MITT y los PP. El ITT refleja lo que ocurre realmente en la práctica clínica, permite obtener del ensayo clínico información más próxima a la efectividad del tratamiento (efecto más cercano a las condiciones reales), en que se dan circunstancias como diagnóstico incorrecto, falta de adhesión al tratamiento, etc . Ello es importante en los ensayos clínicos con una orientación pragmática, y es importante desde nuestro punto de vista , en que nos interesa estimar la efectividad del fármaco cuando lo empleemos en nuestros pacientes.

En muchos ensayos se presentan tablas con ambos tipos de resultados (ITT y PP) pues pueden aportar información complementaria para facilitar la interpretación. Ejemplo en los ensayos de antiretrovirales es muy relevante el grado de cumplimiento terapéutico. En los análisis ITT se estudian todos los pacientes de ambos grupos, con independencia de la adherencia. En el análisis PP, solo aquellos que han seguido de forma correcta el tratamiento. El análisis ITT nos da una estimación global de la eficacia, el análisis PP nos da una estimación de la eficacia de los buenos cumplidores. Si el médico prescribe el tratamiento a un paciente concreto que conoce que es buen adherente, lo esperable es que obtenga un resultado de efectividad similar al análisis PP.

Los análisis que evalúan la seguridad de un fármaco, deben hacerse sobre ITT, ya que el motivo de que un paciente abandone el tratamiento puede estar relacionado con la aparición de una reacción adversa.

En los ensayos clínicos de equivalencia, el análisis por ITT deja de ser conservador, porque permite concluir con más facilidad que los resultados no son estadísticamente significativos y por tanto inferir que las alternativas en estudio son terapéuticamente equivalentes.

Tabla 2

Ejemplo: Definición y número de pacientes evaluados en cada tipo de análisis

Voriconazol en aspergilosis invasora vs Anfotericina B deoxicolato. Herbrecht R, et al. *N Engl J Med* 2002;347:408-15

Análisis por intención de tratar (ITT): Son los reclutados y aleatorizados N=391

Análisis de seguridad: Son los reclutados y aleatorizados y al menos una dosis del medicamento correcto según aleatorización. (Se excluyen 12 pacientes porque no recibieron tratamiento alguno). N= 379

Análisis por intención de tratar modificado (MITT): ITT + Diagnóstico confirmado definitivo probable por un grupo independiente. Se excluyen 102 pacientes por falta de diagnóstico confirmado de aspergilo invasiva. N =277

Análisis CCE (Análisis en pacientes con criterios de evaluación, similar a análisis PP) : Los incluidos en el análisis MITT + No terapia antifúngica concomitante + no violaciones protocolo + adecuada evaluación al final del tratamiento + haber recibido al menos 5 días del tratamiento del estudio. N=242.

Tabla 3

Ejemplos: Resultados expresados por análisis ITT y por análisis PP

ESTUDIO M98-863 Antirretrovirales en pacientes naives.

Estudio fase III, randomizado, doble ciego compara Lopinavir/Ritonavir con Nelfinavir (gold estándar) en combinación con d4T/3TC, en pacientes naives. Se analiza la eficacia en términos de respuesta virológica (proporción de pacientes con <400 copias a la semana 34) y duración hasta la semana 48. n ° de pacientes: 653

	Lopinavir/ Ritonavir	Nelfinavir	p	NNT
PCR-HIV<400 por intención de tratar (ITT)	75%	63%	P<0,001	8,3
PCR-HIV<400 tratados (PP)	93%	82%	P<0,001	9,1

A las 48 semanas el 75% de los pacientes con Lopinavir/Ritonavir vs 63% en Nelfinavir tenían menos de 400 copias. P<0,001.(análisis por intención de tratar). Los resultados si se considera pacientes en tratamiento son 93% y 82% (análisis per protocolo) respectivamente p<0,001.

B) ASPECTOS SECUNDARIOS

Además de los tres puntos mencionados, existen una serie de características secundarias que orientan sobre la calidad del ensayo clínico:

Enmascaramiento. Las expectativas de éxito o fracaso de la intervención, tanto de los pacientes como de los investigadores, pueden influir en la evaluación de la respuesta. Este problema puede prevenirse mediante el uso de 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. La mejor técnica es la del **doble ciego**, según la cual ni los pacientes ni los investigadores conocen el tratamiento asignado. No siempre es posible aplicar estas técnicas. Como mínimo es recomendable que las personas que midan la respuesta desconozcan el tratamiento asignado, de forma que se reduzca el posible sesgo.

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 asegura que se haya producido una distribución equilibrada de todas las variables relevantes, especialmente cuando el número de sujetos 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.

Si los grupos no son similares por alguna variable, el análisis estadístico permitirá controlar la influencia de dicha diferencia y obtener una estimación ajustada del efecto del tratamiento mediante el análisis estratificado o una técnica multivariante.

La comparabilidad de los grupos debe mantenerse a lo largo de todo el seguimiento. Si uno de los grupos ha sido seguido con mayor atención que el otro, existe una mayor probabilidad de detectar cualquier suceso que ocurra durante el seguimiento, e incluso de modificar la intervención o de administrar otros tratamientos, interfiriendo con la evaluación de la respuesta.

VARIABLES INTERMEDIAS O VICARIANTES. Muchos ensayos clínicos cuantifican los resultados de eficacia mediante una variable intermedia. Las variables intermedias deben reflejar el resultado final de forma convincente, de lo contrario la fortaleza de las conclusiones se debilita. La variable intermedia debe presentar una asociación consistente, fuerte e independiente con el desenlace clínico. Ver tabla 4.

Tabla 4

Características de una variable intermedia ideal.

- Fiable, reproducible, clínicamente disponible, fácilmente cuantificable y capaz de mostrar un efecto "dosis-respuesta".
- Verdadera predictora de la enfermedad (o del riesgo de enfermar).
- Su relación con la enfermedad debe ser biológicamente plausible.
- Sensible.
- Específica.
- Punto de corte preciso entre valores normales y anormales.
- Valor predictivo positivo aceptable.
- Valor predictivo negativo aceptable.
- Susceptible de control de calidad.
- Cambios de la variable intermedia deben reflejar rápidamente y con precisión la respuesta al tratamiento, así como normalizar sus valores en las fases de remisión o cura.

Tabla 5

Resumen. Pasos básicos a seguir para valorar la validez interna de un ensayo clínico

(Adaptado de Jimenez 1997, Guyatt 1995)

1. Validez interna del estudio (metodología correcta)	
preguntas principales:	¿La asignación de sujetos ha sido aleatoria ? ¿Han sido seguidos todos los sujetos ? ¿Se ha utilizado la estrategia de análisis por intención de tratar ?
Preguntas secundarias	¿ se ha utilizado alguna técnica de enmascaramiento? ¿ los grupos eran comparables al inicio del estudio ? ¿ los grupos han sido seguidos y tratados de la misma forma ?, excepto el tratamiento de estudio ? ¿Se emplea una variable intermedia o una variable final?. ¿En el primer caso está validada?

2.2 -EFICACIA Y VALIDEZ EXTERNA O APLICABILIDAD: HACIA LA EFECTIVIDAD

Los ensayos clínicos dan resultados de eficacia, pero lo que nos interesa cuando aplicamos sus conclusiones a nuestros pacientes, es conocer cual será su efectividad real, teniendo en cuenta tanto las características de base del grupo de candidatos a tratar, como los aspectos de atención asistencial vigentes en nuestro medio.

Los estudios con criterios **restrictivos** tienen mayores limitaciones para la extrapolación. Los estudios **pragmáticos** son más fácilmente generalizables. Mayor validez externa se asocia al carácter **multicéntrico** del estudio ya que supone la participación de múltiples instituciones sanitarias, de pacientes de diversas procedencias y de diferentes estilos de práctica clínica.

Debe valorarse si 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, etc son adecuadas al problema de salud de interés. Si los pacientes tienen las mismas características que los pacientes incluidos en el estudio, los resultados serán perfectamente aplicables. Por ello los estudios deben indicar claramente los criterios de selección, la procedencia y las características de los sujetos estudiados.

Es conveniente tener en cuenta las siguientes cuestiones sobre la utilidad práctica de los resultados del ensayo. Responder a unas simples preguntas nos ayudan a centrar el tema:

¿Es el fármaco con que se compara, el estándar adecuado y el que se emplea en la práctica habitual de mi hospital? Si, No

¿En caso de que el fármaco de comparación no sea el estándar de mi hospital, aporta el ensayo pivotal evidencia indirectas de utilidad? Si , No.

¿Las características basales de mis pacientes son similares a las del ensayo? Si, No. Ver criterios de inclusión y exclusión.

¿Los resultados pueden aplicarse a mi práctica? Si, No.

¿El seguimiento clínico y asistencial de mis pacientes es similar a las del ensayo? Si No.

-En los ensayos en que se utiliza una variable intermedia, ¿es ésta consistente con el resultado final?

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

Las evidencias de superioridad.

Para considerar un fármaco superior a otro y justificar su inclusión en el hospital, debemos tener en cuenta dos puntos básicos más:

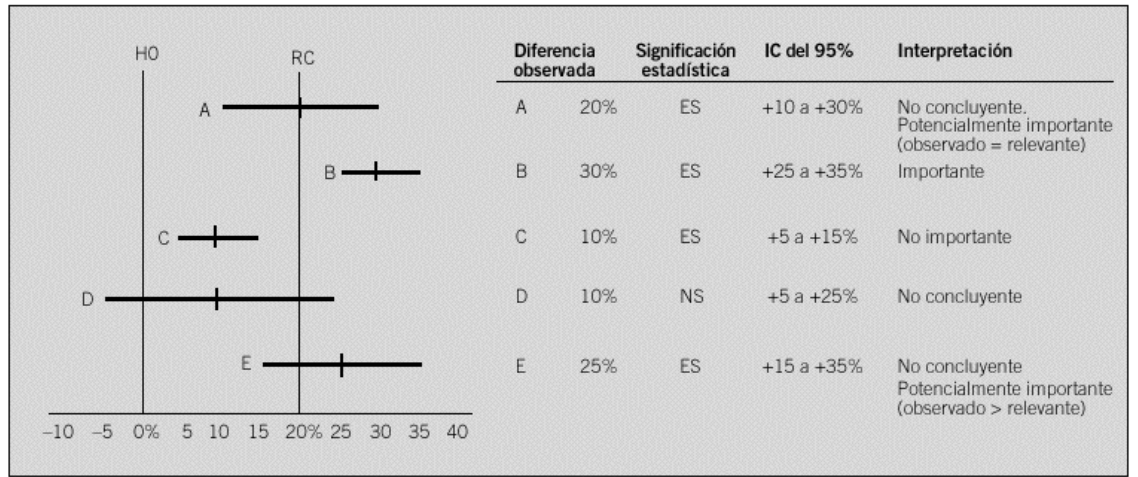
-Valorar si la **magnitud del efecto tratamiento** es de relevancia clínica. Para ello debemos establecer cual es el valor mínimo que consideramos clínicamente relevante. Si el valor encontrado en el ensayo es superior y su IC95% de confianza no incluye dicho valor mínimo, podemos concluir que el tratamiento en estudio es superior. (Ejemplo B de la gráfica). Si el valor es inferior, estadísticamente significativo y su IC95% no incluye el valor mínimo clínicamente relevante, podemos considerar que el medicamento es algo superior pero sin importancia clínica (Caso C de la gráfica). Los casos A y E indican superioridad y de posible importancia clínica , aunque no son concluyentes hay que tenerlos en cuenta como indicativos de superioridad.

Gráfica 1

Resultados hipotéticos de un estudio que compara un tratamiento activo frente a un placebo y utiliza como medidas la diferencia de eficacia (RAR). Se considera que la mínima RAR relevante clínicamente es del 20 %.

ES: Diferencia significativa estadísticamente. NS: No significativa estadísticamente.

Argimon JM Med Clin (Barc) 2002, 118:382-4.



-Los datos de subgrupos y el análisis pormenorizado del ensayo son de gran interés para determinar al final las **condiciones de uso** del nuevo fármaco. Ello ayuda a delimitar los subgrupos de pacientes que más se pueden beneficiar de los efectos del medicamento y establecer las condiciones de uso en el hospital. De hecho la mayor parte de los nuevos medicamentos, se introducen por su superioridad en pequeños grupos de pacientes y condiciones de uso estrictas.. Cuando un estudio presenta análisis de subgrupos hay que ser muy cauteloso y evaluar si el subgrupo es grande, si responde a una hipótesis establecida previamente al inicio del estudio, si se han analizado pocos subgrupos y si se han observado los mismos resultados en otros estudios. Si no se cumplen estas circunstancias habrá que ser escéptico.

Las evidencias de equivalencia terapéutica

La decisión de incluir un nuevo medicamento como equivalente terapéutico, se realiza según las evidencias. Diferentes tipos de estudios nos aportan información para la toma de decisiones: (**Ver tema C del capítulo III información ampliada**).

- Estudios pragmáticos, sobre todo de tipo observacional
- Ensayos directos entre dos opciones con el objetivo de demostrar equivalencia o no inferioridad
- Estudios directos entre dos opciones con el objetivo de demostrar superioridad pero no aparecen diferencias estadísticas significativas entre las opciones

-Evidencias indirectas por la existencia de ensayos con un tercer comparador

Tabla 6 Resumen. Aspectos básicos validez externa o aplicabilidad
Relevancia de los resultados
¿ se presentan diferencias significativas?
¿ Se presentan los IC de los resultados principales y subgrupos?
¿ cual es la magnitud del efecto tratamiento y si es de relevancia clínica ?
¿en caso de utilizar una variable intermedia, es esta consistente con el resultado final ?
Utilidad práctica de los resultados
¿el fármaco con que se compara es el estandar adecuado y el que se emplea en la práctica habitual de mi hospital?
¿ las características basales de mis pacientes son similares a las del ensayo? Ver criterios de inclusión y exclusión.
¿ los resultados pueden aplicarse a mi práctica ?
¿ el seguimiento clínico y asistencial de mis pacientes es similar a las del ensayo ?
¿ se han tenido en cuenta todos los desenlaces clínicamente importantes ?

2-3. RELEVANCIA DE LOS RESULTADOS

La decisión de si un tratamiento es o no más eficaz que el de referencia no debe basarse exclusivamente en el resultado de una prueba de significación estadística. Si el lector solo conoce el valor de p, sabrá si el estudio ha demostrado o no que existe un efecto, pero no si este es importante en clínica. Es fundamental que el artículo proporcione una estimación de la **magnitud** del efecto del tratamiento para que el lector pueda evaluar si es un efecto de importancia clínica, y una estimación de la **precisión** con sus intervalos de confianza.

A-FORMA DE PRESENTAR LA MAGNITUD DEL EFECTO.

En la mayoría de los estudios pivotaes se emplean variables binarias de tipo "si" o "no", para la obtención de los resultado principales (ejemplos: % de muerte de pacientes, % curación de pacientes, % de pacientes con determinada respuesta clínica). Muchas variables continuas es habitual transformarlas y presentarlas en binarias en la publicación original.

La forma de expresar los resultados puede ser diferente y es necesario familiarizarse con los distintos índices y sus relaciones. Muchas veces los resultados se expresan en medidas relativas, sobre todo la Reducción del

Riesgo Relativo (RRR), ya que de esta forma se obtienen valores más espectaculares y aparentes. También con frecuencia se usan el Riesgo Relativo (RR) y el Odds Ratio (OR) que se han incorporado como elemento de expresión de resultados no solo en los estudios observacionales, sino también en los ensayos clínicos y en los meta-análisis.

Las medidas absolutas se basan en calcular la diferencia entre frecuencias del resultado observado en cada grupo de sujetos, mientras que las medidas relativas se basan en el cociente entre dichas frecuencias, cuantificando cuántas veces es más probable que ocurra un suceso en un grupo que en otro.

Las formas habituales de presentación de resultados son:

Medidas absolutas:

- Reducción Absoluta del Riesgo (RAR)
- Número Necesario a Tratar (NNT)

Medidas relativas:

- Reducción Relativa del Riesgo (RRR)
- Riesgo relativo del grupo tratamiento respecto al grupo control (RR)
- Odds ratio (OR)

Ver ejemplos de concepto y cálculo en la tabla

Tabla 7		
EJEMPLO. Forma de presentar los datos de eficacia de un ensayo clínico.		
<i>Ejemplo sobre resultados de curaciones en un ensayo clínico en que el Grupo A corresponde al grupo control y Grupo B al grupo intervención.</i>		
<i>Fórmula</i>	<i>Terminología</i>	<i>Cálculo y resultado</i>
A	Riesgo (incidencia, proporción) grupo control o riesgo base	Ejemplo 15/100 = 15 % = 0,15
B	Riesgo (incidencia, proporción) grupo intervención	Ejemplo 20/100 = 20 % = 0,2
B – A	RAR (Reducción Absoluta del Riesgo o Diferencias de Riesgo o Riesgo atribuible)	0,2-0,15 = 0,05
1 / (B-A)	NNT (Numero Necesario a Tratar). Es decir que es igual a 1/RAR	1/(0,2-0,15) = 1/0,05 = 20
B / A	RR (Riesgo Relativo o Razón de Riesgos)	0,20 / 0,15 = 1,33
B' / A'	OR (Odds ratio)	20/80: 15/85 = 0,25/0,176 = 1,42
[(B-A) / A] x 100	RRR (Reducción Relativa del Riesgo). Es decir que es igual a 1-RR	[(0,2-0,15) / 0,15] x 100 = 33 % = 0,33
Odds de B = B' = casos/no casos grupo intervención: 20/80 = 0,25		
Odds de A = A' = casos/no casos grupo control: 15/85 = 0,176		
<i>Nota: En este ejemplo no se tiene en cuenta el signo + ó - de la diferencia (A-B) o de la diferencia (B-A). Lo importante es referirlo a la variable del estudio (mortalidad, curaciones, etc) y comprender si aumenta o disminuye el riesgo respecto al grupo control (ver ejemplo tabla 12 del capítulo 2 referido a mortalidad) . En algunos casos prácticos puede ser conveniente expresar el signo positivo o negativo, sobre todo cuando hay resultados parciales en sentidos divergentes (ver el ejemplo de la tabla 14 del capítulo 2).</i>		

Para nuestro propósito conviene manejar las medidas absolutas.

Medidas absolutas:

-Concepto y cálculo de RAR

El RAR (Reducción Absoluta del Riesgo o Diferencias de Riesgo o Riesgo atribuible) es la forma más simple de expresar la diferencia de eficacia entre los grupos estudiados. Es la diferencia de proporciones de un suceso entre el grupo control y el grupo intervención.

Los datos básicos de proporción o incidencia del evento de cada grupo estudiado y su diferencia o RAR nos informan de un forma simple de la magnitud de la diferencia de eficacia.

La RAR debe conocerse antes de tomar decisiones sobre la ventaja de un fármaco y sobre que pacientes deben tratarse.

-Concepto y cálculo de NNT

El NNT es el número necesario de pacientes a tratar para producir una unidad adicional de eficacia. El NNT, es un dato fácil de obtener de los resultados de los ensayos clínicos, siempre y cuando la medida de eficacia venga expresada en forma de una variable binaria.

El NNT nos es útil para dimensionar la eficacia del fármaco mediante la expresión del grado de “esfuerzo” a realizar para conseguir una unidad adicional de eficacia. También nos será muy útil para los cálculos farmacoeconómicos.

El $NNT = 1/RAR$. Si el RAR lo expresamos en porcentaje, entonces $NNT = 100/RAR$

Tabla 8.

Cálculo del Numero necesario de pacientes a tratar (NNT)

El NNT es el inverso de la reducción del riesgo absoluto (1/RAR) y se puede calcular a partir de los resultados de eficacia de un ensayo clínico, aplicando una sencilla fórmula

<p>Numero Necesario a Tratar</p> $NNT = \frac{100}{b-a}$	<p>Resultados de un ensayo clínico: <i>a: Porcentaje de pacientes del grupo tratado con placebo o con el fármaco de referencia con resultados de eficacia (brazo control). b: Porcentaje de pacientes del grupo tratado con el nuevo fármaco con resultados de eficacia (brazo experimental). b-a: Reducción del riesgo absoluto (RAR) expresado en porcentaje.</i></p>
---	---

Medidas relativas

La expresión de los resultados exclusivamente como medidas relativas, hecho que ocurre a menudo al difundir las conclusiones de un ensayo clínico, puede resultar tendenciosa y hacer que se sobrevalore el efecto beneficioso real de un tratamiento. El problema está en que las medidas relativas no tienen en cuenta el riesgo basal o de partida de los pacientes, y por ello no permiten diferenciar los grandes beneficios de los pequeños. Ver tablas 9 y 10

-Concepto de RRR. Alerta con la forma de presentar los datos de eficacia:

El RRR es la diferencia en las proporción de acontecimientos entre el grupo control y el de tratamiento, dividido por la proporción de acontecimientos en el grupo control.

$$[(B-A) / A] \times 100$$

Tabla 9

Ejemplo de RAR y RRR

Si un fármaco disminuye la mortalidad de un 6 % a un 3 %, podemos decir realmente que :
-Ha disminuido la mortalidad un 3 % en términos absolutos (6% -> 3%) (RAR).
-Ha disminuido la mortalidad un 50 % en términos relativos (6% -> 3%) (RRR).
-Ha aumentado la supervivencia del 94 % al 97%, es decir un 3,2 % en términos relativos (RRR)
En todos los casos decimos la verdad
Por ello es mejor hablar en términos de RAR (Reducción absoluto del riesgo), que todos entendemos de forma más clara y precisa.

Tabla 10

Medidas del efecto de un tratamiento en función del riesgo basal de los sujetos del grupo control, suponiendo riesgos basales de 40%, 20%, 10 % y 5 %.

	Riego basal del grupo control			
	40 %	20 %	10 %	5 %
Incidencia en el grupo tratado	30 %	15 %	7,5 %	3,75 %
RRR (Reduccion Relativa del Riesgo)	25 %	25 %	25 %	25 %
RAR (Reducción Absoluta del Riesgo)	10 %	5 %	2,5 %	1,25 %
NNT (Numero Necesario a Tratar)	10	20	40	80

-Concepto y cálculo de RR

El RR (Riesgo Relativo) es una relación entre la proporción de eventos del grupo intervención y del grupo de referencia o control.

Es una medida de los resultados relativa relacionada directamente con el RRR (Reducción del Riesgo Relativo), de hecho $RRR=(1-RR) \times 100$.

Por ejemplo la proporción de pacientes que sufren una reacción adversa con el tratamiento estandar es 0,10 (10 %), mientras que la proporción de pacientes que presentan la misma reacción adversa con el tratamiento en estudio es 0,125 (12,5%). EL RR es igual a 1,25 (es decir el cociente de 0,125 / 0,1). También podemos decir que con el tratamiento estandar un 25 % menos de pacientes sufren la reacción adversa [$RRR=(1-1,25) \times 100=25\%$].

Otro ejemplo: La mortalidad en el brazo placebo es del 15%. Este porcentaje es del 5% en el brazo intervención. EL RR es 0,33 (cociente de 0,05/0,15). Es decir que hay un 67% más de mortalidad en el grupo placebo [RRR=(1-0,33) x 100=67%].

RR toma valores entre 0 y infinito. El valor 1 es el valor neutro y significa que no hay diferencias entre los grupos comparados. Cuando nos acercamos al valor 0 o a un valor muy superior a 1, significa que la diferencia entre grupos es muy grande.

Si toma un valor >1 significa que el grupo cuya resultado hemos puesto en el numerador tiene una mayor proporción que el que hemos puesto en el denominador, lo contrario si el RR es < 1.

(Nota: Si intercambiamos el lugar donde ponemos las dos proporciones el RR será el inverso (1/RR)).

-Concepto y cálculo OR (Odds Ratio o Momio Ratio)

La medida del efecto más empleada en metaanálisis, tanto para expresar los datos de cada ensayo como el resultado global comparado es la OR (Ejemplo Biblioteca Cochrane). Estadísticamente se prefiere la OR a la RR debido fundamentalmente a sus propiedades matemáticas para el cálculo del resultado global. Permite mediante la regresión logística, ajustar la relación en estudio por el efecto de otras variables. Actualmente en muchos ensayos clínicos los resultados se expresan en OR.

La Odds. Es la razón en la que el numerador es la probabilidad de que ocurra un suceso y el denominador es la probabilidad que tal suceso no ocurra. En las apuestas de juego de pelota de los frontones se le llama "momio". Habitualmente se estima mediante el número de pacientes que cumplen los criterios de un objetivo dividido por el número de pacientes que no cumplen el objetivo.

La OR (Odds Ratio). La división de una *odds* por otra *odds* es una *razón de odds*.

La fórmula de cálculo es:

$$OR = p1 / (1-p1) : p2 / (1-p2).$$

Ejemplo, si tras un tratamiento antibiótico en un grupo de 10 pacientes, 4 presentan diarrea (4/6=0,66) y en el grupo control de 10 pacientes 1 presenta diarrea (1/90=0,11), el OR será de = 0,66/0,11 = 6.

En una tabla 2x2.la OR también se puede calcular según la fórmula:

$$OR= AD/BC$$

	Éxito	Fracaso
Grupo 1 (intervención)	A	B
Grupo 2 (control)	C	D

-Relación entre RR y OR.

En eventos que se produzcan con baja incidencia (<10 %), el valor numérico del OR y el del RR serán semejantes. Cuando el suceso es más frecuente del 10 % pero la diferencia entre grupo control y grupo intervención es pequeña también resultan similares las estimaciones del RR y de la OR.

Tabla 11

Ejemplos: Expresión de resultados como RR y OR

Ejemplo 1, baja incidencia (<10%):

Riesgo en grupo control= 0,05.

Riesgo en grupo intervención= 0,1

RR=0,1 / 0,05 = 2

Odds grupo control = 0,05/0,95 = 0,052

Odds grupo intervención:0,10/0,90 = 0,111

OR= 0,111 / 0,052 = 2,13

Ejemplo 2, alta incidencia (>10%) y diferencia entre grupo control e intervención grande:

Riesgo en grupo control=0,4.

Riesgo en grupo intervención= 0,9

RR=0,9/0,4 = 2,25

Odds grupo intervención0,9/0,1= 9

Odds grupo control =: 0,4/0,6 = 0,66

OR= 9,0 / 0,66 = 13,5

Las RR es un cociente o relación de *proporciones* y la OR es un cociente o relación de dos *odds*. Tanto las proporciones como las odds expresan lo mismo , pero usan dos escalas numéricas distintas, las proporciones oscilan entre 0 y 1 y las odds entre 0 e infinito. Para pasar de una escala a otra se utilizan las expresiones : odds=p/(1-p); y p=odds / (1+odds).

En general los resultados expresados en OR serán mucho más aparentes es decir con valores más extremos o alejados de 1 (más cerca de 0 o más cerca de infinito) que los expresados en RR, especialmente cuando la incidencia de suceso de alguno de los dos grupos es superior al 10 % y las diferencias entre ellos son grandes. Esta circunstancia es frecuente en los ensayos clínicos. Los IC de los OR también suelen ser más amplios que los de RR. En los estudios epidemiológicos se suelen manejar tasas de incidencia muy bajas, y por ello los valores de OR y RR son similares.

-Relación entre medidas absolutas y relativas

Un RR o un OR pueden ser difíciles de traducir en beneficio si no se traduce en medidas de la diferencia absoluta de riesgos. El RAR y el NNT se pueden calcular a partir del RR conociendo la incidencia del

evento en el grupo control. También existen fórmulas para calcular el NNT derivado de los resultados de un ensayo clínico a partir de la OR.

En internet existen diferentes programas de acceso libre que facilitan su cálculo rápido. Ver cuadro.

Tabla 12

Cálculo de NNT, RR, OR y RRA con intervalos de confianza 95 %, programas en internet

“Risk Reduction Calculator” <http://araw.mede.uic.edu/cgi-bin/nntcalc.pl>
(gratuito)

“EBM calculator” <http://www.cebm.utoronto.ca/palm/ebmcalc/ebm.zip>
(gratuito). Es un programa para agendas PDA tipo PALM

“UBC Clinical Significance Calculator”
<http://www.healthcare.ubc.ca/calc/clinsig.html>

“Calculador sisa” <http://home.clara.net/sisa/two2hlp.htm>

Otros programas estadísticos de acceso libre en internet:

http://www.fisterra.com/material/investiga/program_internet/program_internet.htm

Nota: Para calcular el IC, debe conocerse: - el n de cada grupo y los resultados de eventos (+/-) cada grupo para hacer la tabla de 2 x 2

También hay tablas

Tabla 13.

Cálculo del NNT a partir de una OR. para un ensayo clínico (Mc Lister JAMA 2000)

Incidencia del evento en el grupo control	Intervención terapéutica (OR)				
	0,50	0,60	0,70	0,80	0,90
	NNT				
0,05	41	52	69	104	209
0,1	21	27	36	54	110
0,2	11	14	20	30	61
0,3	8	10	14	22	46
0,5	5	8	11	18	38
0,7	6	9	13	20	44
0,9	12	18	27	46	101

Existe evidencia empírica que sugiere que los índices relativos OR y RR tienden a ser más parecidos entre diferentes ensayos del mismo tratamiento que los índices absolutos RAR y NNT.

Estos últimos dependerían del tratamiento y también del riesgo basal. Así en determinados ensayos puede pasar que el subgrupo de pacientes con un riesgo base de mayor gravedad se beneficien más del tratamiento (ejemplo Drotrecogina alfa en sepsis), mientras que en otros

casos ocurre lo inverso, el subgrupo de pacientes con una mejor situación de base se "curarán" más (Ejemplo: medicamento para tratamiento de una infección).

Un aspecto importante, para decidir si el resultado de un ensayo, expresado en RAR o NNT es aplicable a nuestro medio, es evaluar si el riesgo basal de los pacientes del ensayo o de algunos de los subgrupos es similar al de nuestro medio.

La aplicación de los índices relativos a situaciones basales diferentes al del ensayo constituye una aproximación, que a pesar de las evidencias empíricas mencionadas, debe realizarse con cautela.

B) PRECISIÓN CON QUE SE ESTIMA EL EFECTO

Toda expresión de la magnitud del efecto debe acompañarse de sus intervalos de confianza. El verdadero efecto de un tratamiento es desconocido. La medida observada no es más que una estimación puntual del mismo. Si se hubiera estudiado una muestra de sujetos diferentes, se habría obtenido una estimación probablemente similar , pero no exactamente igual. Se suele realizar el cálculo del llamado intervalo de confianza del 95 %. Obviamente la precisión expresada por la amplitud de dicho intervalo está muy relacionada con el tamaño de la muestra. Cuanto mayor es el numero de sujetos, mayor es la precisión y más estrecho el intervalo de confianza.

La precisión se estima mediante el intervalo de confianza (IC) del 95 %, que es aquel intervalo entre cuyos límites se tiene el 95 % de confianza de que se encuentre la verdadera magnitud del efecto. EL IC presenta grandes ventajas respecto al grado de significación estadística.

Cuando se emplea como medida del efecto una diferencia como RAR o RRR, si el IC del 95 % incluye el valor 0, se concluirá que el resultado no es estadísticamente significativo. Si por el contrario excluye el valor 0, se concluirá que la diferencia es estadísticamente significativa.

Cuando el IC se aplica a la estimación de un RR o una OR , no se presentan diferencias estadísticamente significativas cuando el IC 95% incluye el valor 1.

Además de saber si el resultado es o no estadísticamente significativo, el IC permite conocer entre que límites es probable que se encuentre la verdadera magnitud, y ver si incluye o excluye el mínimo valor considerado de relevancia clínica.

La mayoría de las publicaciones incorporan actualmente los IC95%, pero no siempre. Para el calculo de los intervalos de confianza del NNT y las otras medidas, hay programas que permiten ayudarnos. Ver tabla anterior con programas de cálculo disponibles en internet.

3. DECIDIR EN LA PRACTICA

¿Son válidos y aplicables los resultados?

En resumen al revisar un ensayo debemos incidir en los aspectos básicos de validez interna, sobre todo si el tipo de análisis es por intención de tratar o per protocolo y en este caso la proporción de pacientes no evaluados.

También es importante la variable empleada en los resultados, ver su relevancia clínica y si se trata de una variable intermedia, que sea realmente predictiva de resultados finales de peso clínico.

Conviene evaluar los resultados en variables absolutas y ver si la magnitud de la superioridad es relevante, tanto en el resultado principal como en los subgrupos. Los IC95% deben conocerse para definir esta superioridad.

Los resultados principales y secundarios deben estar definidos previamente como objetivos del estudio y estar expuestos en el método.

En cuanto a la validez externa, lo más importante es que medicamento de referencia sea el estándar adecuado, junto con las demás cuestiones del apartado 2.2.

La cuestión es que muchas veces la información y evidencia aportada por el ensayo es incompleta y genera incertidumbre sobre su interés práctico en nuestros pacientes.

Para facilitar de forma sistemática y estructurada la evaluación de la calidad de un ensayo, se han confeccionado listas-guía o check-list, que pueden ser de tipo cualitativo o mediante escalas cuantitativas. Se han publicado multitud de listas-guía, más o menos complejas (ver en anexo la lista CONSORT) sirven para repasar de forma sistemática cada uno de los puntos importantes.

Desde un punto de vista prácticos puede ser preferible emplear algunos conceptos de la MBE, como son las escalas de clasificación de la evidencia científica que diferencian de forma jerárquica los distintos niveles de la evidencia en función del rigor científico del diseño del estudio. De aquí se derivan la magnitud y rigor de las recomendaciones que se traducen en una definición del grado de evidencia y recomendación.

Hay diferentes escalas de gradación de la calidad, bastante similares entre sí, aunque difieren en algunos matices. En una revisión reciente (*Martínez-López I et al Informes normalizados de evaluación de medicamentos: resultados de dos años de aplicación. 2003 Aten Farm (en prensa)*), nosotros hemos aplicado una escala de evidencia científica modificada de *Jovell et al (Jovell AJ, Navarro-Rubio MD: Evaluación de la evidencia científica. Med Clin (Barc) 1995; 105:740-3)*, que nos ha permitido clasificar los ensayos clínicos pivotaes de los últimos medicamentos estudiados. Es sólo una aproximación, que se basa en

establecer un nivel y restar categorías en función de las condiciones de validez interna o externa no adecuadas. De esta forma podemos resumir en una frase el nivel de evidencia de cada estudio.

Tabla 14
Niveles de evidencia.
Modificado de Jovell AJ , Navarro-Rubio MD: Evaluación de la evidencia científica. Med Clin (Barc) 1995; 105:740-3),

	Nivel de evidencia	Diseño	Condiciones rigurosidad científica	Validex externa o aplicabilidad
I.	Excelente	Metanálisis de ensayos controlados y aleatorizados	Calidad de los estudios primarios	Compara con el estándar correcto
II	Adecuada	Ensayo controlado y aleatorizado de muestra grande (n > 500 para este chec-list)	Poder estadístico Calidad del estudio, pérdidas seguimiento, no análisis post-hoc, Variable principal	Compara con el estándar correcto
III	Regular	Ensayo controlado y aleatorizado de muestra pequeña (n < 500 para este chec-list)	Poder estadístico. Calidad del estudio, pérdidas seguimiento, no análisis post-hoc, Variable principal	Compara con el estándar correcto
IV	Baja	Ensayo prospectivo aleatorizado no ciego Ensayo prospectivo controlado no aleatorizado	Controles coincidentes en tiempo. Calidad del estudio, pérdidas seguimiento, no análisis post-hoc, Variable principal	Compara con el estándar correcto
V	Baja	Ensayo retrospectivo no aleatorizado		
VI	Baja	Estudio de cohortes Estudio de series observacional		
VII	Pobre	Casos y controles		
VIII	Pobre	Series clínicas, Estudios descriptivos		

Si el estándar con que se compara no es adecuado (-2 categoría), estudio no publicado (-1 categoría). No calidad. (-1 categoría)

En nuestro caso, durante el periodo de estudio de dos años, se han realizado 28 informes de evaluación. Los niveles de evidencia aportados por los estudios fueron. Nivel I (excelente): 0 casos, Nivel II (adecuada): 10 casos (50%). Nivel III (regular): 4 casos (20%). Niveles IV a VI (baja): 6 casos (30%).

Sin embargo, al margen de clasificaciones que facilitan colocar una “etiqueta”, lo importante es la lectura crítica y pormenorizada del ensayo, que es el aspecto clave para decidir sobre la selección y definir las condiciones de uso de un nuevo medicamento

4. Anexo tema A del capítulo III

La lista Guía Consort

La propuesta CONSORT, realizada por un grupo de expertos, expone las recomendaciones básicas que deben seguirse para publicar ensayos clínicos en revistas biomédicas. Contiene un diagrama de flujo, que sirve de modelo para que los autores de un ensayo expongan de forma ordenada el número de pacientes inicialmente incluidos en el estudio y las pérdidas ocurridas en cada fase junto a los motivos. (Ref: *Altman DG et al: Ann Intern Med 2001; 134:663-694*).

En el congreso de la SEFH del año 2002 de Barcelona, nuestro servicio presentó una comunicación en la que se revisaba la calidad de los ensayos clínicos "pivotal" que han tenido impacto en la introducción de nuevos medicamentos en el hospital en los últimos años. Se evaluaron 8 ensayos clínicos de los fármacos: Darbepoetina, EPO 40.000 UI, Levonorgestrel DIU, Peg-interferon alfa-2b, Tenecteplasa, Tobramicina TNS, Verteporfino y Zoledronato, y se valoró el porcentaje de cumplimiento de los criterios CONSORT. Los criterios se cumplen en un 68% de media (rango 52-96%). (Ref *F. Puigventós et al Comunicación al Congreso de la SEFH. Barcelona 2002*)

Resumen

Tabla 15
ENSAYOS CLÍNICOS CON IMPACTO EN LA INTRODUCCIÓN DE NUEVOS FÁRMACOS EN EL HOSPITAL: APLICACIÓN DE LA LISTA-GUÍA CONSORT
F. Puigventós Latorre, P Ventayol Bosch, I Martínez López, M Crespi Monjo, A Escrivá Torralva, F. Comas Gallardo, O. Delgado Sánchez.

Objetivo: Evaluar la calidad de los principales ensayos clínicos que han justificado la inclusión de los últimos fármacos en el hospital mediante aplicación de la lista-guía CONSORT.

Métodos: Se identifican los ensayos clínicos publicados de los últimos fármacos incluidos en el hospital. Dos revisores de forma independiente evaluaron los artículos, valorando el grado de cumplimiento de los criterios y apartados de la lista-guía CONSORT; en caso de discrepancia fue revisado y consensuado por un tercer revisor.

Resultados: Se han evaluado 8 ensayos clínicos de los fármacos: Darbepoetina, EPO 40.000 UI, Levonorgestrel DIU, Peg-interferon alfa-2b, Tenecteplasa, Tobramicina TNS, Verteporfino y Zoledronato, y se ha valorado el porcentaje de cumplimiento de los criterios CONSORT:

Darbepoetina	EPO-40000	Levonorgestrel	Peg-INF	Tobramicina	TNK	Verteporfino	Zoledrónico
Kidney Intern 2001;60:741-7.	Ann Inter Med 2000; 133:845-54.	BMJ 1998; 316:1122-6.	Lancet 2001; 358:958-65.	N Eng J Med 1999; 340:23-30.	Lancet 1999; 354:716-22.	Arch Ophthalmol 1999; 117: 139-45.	J Clin Oncol 2001, 19: 6558-67.
52 %	83 %	61 %	70 %	54 %	71 %	96 %	59 %

Los criterios se cumplen en una media del 68 % (52-96%), existiendo homogeneidad entre los revisores en un 88 % (sólo el 12 % requirió una tercera revisión). Los criterios que menos se cumplen son: Métodos usados para garantizar calidad de la medida de los resultados (12%), descripción detallada de como se aplica la aleatorización (12%), evaluación del proceso y éxito de cegar (20%), criterios de suspensión del ensayo (33%), y presentación de resultados con intervalos de confianza (50%).

Conclusiones: La mayoría de ensayos evaluados dejan de cumplir una elevada proporción de los criterios. La guía CONSORT es fácil de aplicar y ayuda a sistematizar las revisiones de los ensayos clínicos. Para su aplicación a la evaluación de nuevos medicamentos es necesario valorar otros aspectos adicionales, como la adecuación del grupo control, la relevancia clínica de los resultados y realizar una ponderación de los criterios.

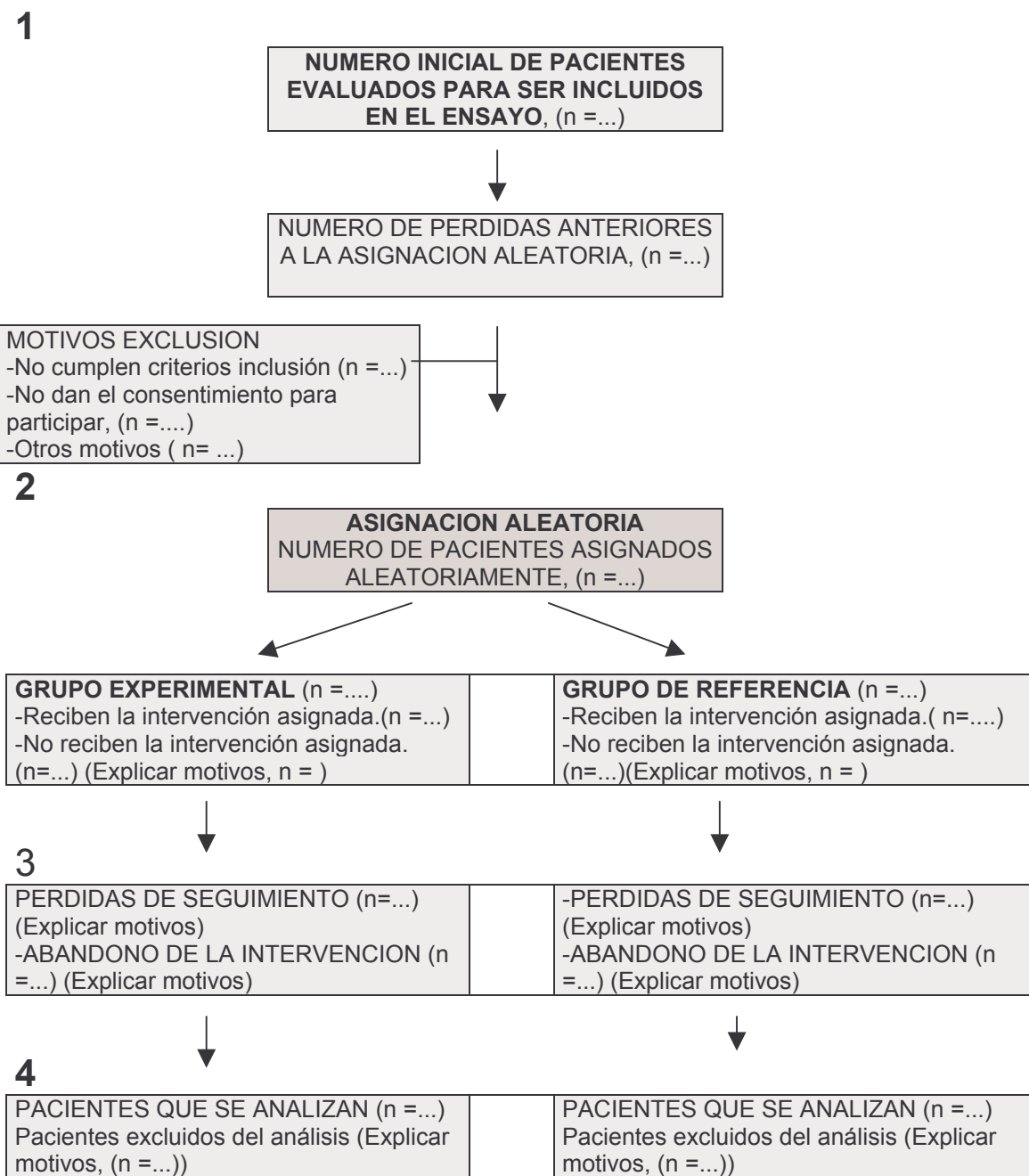
Manual para la evaluación de nuevos fármacos en el hospital
Versión nº 4. Abril 2004

CONSORT 2001			Si	No	No apl
Sección	Item	Descripción. Debe constar:			
TITULO Y RESUMEN					
	1	¿ Cómo fueron los participantes asignados a la intervención ? (ej: "adjudicación aleatoria", "aleatorizados", "asignados aleatoriamente")			
INTRODUCCION					
Antecedentes	2	Antecedentes científicos y explicación racional			
METODOS					
Participantes	3a 3b	Criterios de elegibilidad de los participantes y lugar de recogida los datos			
Intervenciones	4	Detalles precisos de la intervención a realizar en cada grupo y como y cuando será aplicada			
Objetivos	5	Objetivos específicos e hipótesis			
Resultados	6a 6b	Definición clara de las medidas de los resultados primarios y de las medidas de los resultados secundarios, y cuando sea aplicable métodos usados para garantizar la calidad de las medidas (Ej: observaciones múltiples, formación de los asesores).			
Tamaño de la muestra	7a 7b	Cómo fue calculado el tamaño de la muestra y, cuando sean aplicables, exposición de análisis intermedios y criterios de suspensión del ensayo			
Aleatorización: generación de la secuencia	8a 8b	Método usado para generar la secuencia de aleatorización , incluyendo detalles de cualquier restricción (Ej :bloques, estratificación, ...)			
Aleatorización: asignación oculta	9	Método usado para ocultar la asignación aleatoria. (Ej: sobres numerados, teléfono central), especificando si la secuencia es oculta hasta que la intervención es asignada.			
Aleatorización: aplicación	10	Quién genera la secuencia de la asignación, quién incorpora a los participantes, y quién asigna los participantes a sus grupos.			
Ciego (enmascaramiento)	11a 11b	Si aquellos que administran las intervenciones y evalúan los resultados son ciegos respecto a la asignación de grupos. Si ello es así cómo se evaluó el proceso y éxito de cegar.			
Los métodos estadísticos	12a 12b	Métodos estadísticos usados para comparar grupos en los resultados primarios. Métodos para los análisis adicionales, tales como análisis del subgrupo y análisis ajustados.			
RESULTADOS					
Flujo de los participantes	13a 13b	Flujo de los participantes a través de cada etapa (se recomienda un diagrama). Específicamente, para cada grupo estudiado, comunicar el número de participantes asignados aleatoriamente, los que han recibido el tratamiento previsto, los que han terminado el protocolo del estudio, y los analizados para el resultado primario. Descripción de las desviaciones del protocolo del estudio previsto, junto con las razones de dichas desviaciones.			
El reclutamiento	14	Fechas que definen los períodos de reclutamiento y seguimiento.			
Datos base	15	Características demográficas y clínicas cada grupo.			
Números analizados	16	Los números de participantes (denominador) en cada grupo incluidos en cada análisis y especificar si el análisis se ha realizado "por intención de tratar". Indique los resultados en números absolutos cuando es factible (Ej: 10/20, no 50%).			
Resultados y estimación	17	Para cada resultado primario y secundario, un resumen de los resultados para cada grupo, y del tamaño estimado del efecto y su precisión (Ej:., intervalo de la confianza del 95%).			
Análisis auxiliares	18	Comunicar cualquier otro análisis realizado, incluyendo análisis de subgrupo y análisis ajustados, indicando aquellos que se han pre-especificado y aquellos que son exploratorios.			
Efectos adversos	19	Todos los acontecimientos adversos importantes o efectos secundarios en cada grupo de la intervención.			
DISCUSIÓN					
Interpretación	20	Interpretación de los resultados, considerando la hipótesis del estudio, fuentes de sesgos potenciales o de imprecisión, así como los peligros asociados con la multiplicidad de análisis y resultados.			
Generalización	21	Generalización (validez externa) de los resultados de ensayo.			
Evidencia	22	Interpretación general de los resultados en el contexto de la			

general	evidencia actual.			
---------	-------------------	--	--	--

DIAGRAMA CONSORT 2001
 Flujo de las diferentes fases de un EC

1-Reclutamiento
 2-Asignación
 3-Seguimiento
 4-Análisis



Referencias:

Guyatt G, Rennie D, Haywart RSA. *Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice*. AMA ed 2002 <http://pubs.ama-assn.org/misc/usersguides.dtl>

Argimon JM. La ausencia de significación estadística en un ensayo clínico no significa equivalencia terapéutica. *Med Clin (Barc)* 2002; 118 (18): 701-3

Argimon JM. El intervalo de confianza, algo más que un valor de significación estadística. *Med Clin (Barc)* 2002; 118(10): 382-4

Altimiras J et al: *Farmacoepidemiología y estudios de utilización de medicamentos*. EN J. Bonal et al ed *Farmacia Hospitalaria*, 3 ed 2002

Guyatt G et al: *Therapy and understanding results*. in *Users Guides to the medical literature*. AMA ed 2002

Bucher H et al : *Therapy and applying the results*. In *Users Guides to the medical literature*. AMA ed 2002

Oxman A et al. *Summarizing the evidence*. In *Users Guides to the medical literature*. AMA ed 2002

Abraira V *Medidas del efecto de un tratamiento: Reducción absoluta del riesgo, reducción relativa del riesgo y riesgo relativo*. *SEMERGEN* 2000, 26: 535-6

Abraira V. *Medidas del efecto de un tratamiento: odds ratio y numero necesario para tratar*. *SEMERGEN* 2001, 27:418-20

Abraira V. *Estimación: intervalos de confianza*. *SEMERGEN* 2002; 28(2): 84-5

Abraira V: *¿Qué es el análisis por intención de tratar?* *SEMERGEN* 2000; 26: 393-4

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: 192-5

Martínez-gonzález MA et al: *¿Qué es una odds ratio?*. *Med Clin (Barc)* 1999; 112:416-22

Jiménez Villa J: *Lectura crítica de la literatura científica (I) validez del estudio*. *FMC Formacion Medica Continuada en Atención Primaria*. 2000, 7, 4: 209-220

Jimenez Villa J. *Lectura crítica de la literatura científica II. Evaluación de los resultados*. *FMC Formacion Medica Continuada en Atención Primaria*. 2000, 7, 5: 283-296

Jovell AJ, Navarro-Rubio MD: *Evaluación de la evidencia científica*. *Med Clin (Barc)* 1995; 105: 740-43

Granados A: *LA evaluación de las tecnologías médicas*. *Med Clin(Barc)* 1995; 104: 581-585.

Ioannidis JPA, Lau J: *Can quality of clinical trials and meta-anayses be quantified ?* *Lancet* 1998; 352: 590-91

Berlin JA, Rennie D: *Measuring the quality of trials. The quality of quality scales*. *JAMA* 1999, 282, 11: 1083-84.

Juni P et al: *The hazards of scoring the quality of clinical trials for meta-analysis*. *JAMA* 1999; 282: 1054-60.

Barton S: Which clinical studies provide the best evidence. BMJ 2000; 321: 255-56

Jiménez Villa J: Lectura crítica de un ensayo clínico. Cuadernos de gestión: 1997, 3, 2: 76-89

García Lopez 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

The Asilomar Working Group on Recommendations for Reporting of Clinical Trials in the Biomedical Literature: Checklist of information for inclusion in reports of clinical trials; Ann Inter Med 1996, 124, 8: 741-743

Guyatt GH, Sackett DL, Cook DJ: User's guides to the medical literature. II How to use and 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 and 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.

Mora R. Como mejorar la calidad estadística de los artículos presentados a revistas biomédicas: lista de comprobación para los autores. Med Clin (Barc) 1999; 113: 138-149

Puigventós F. Criterios para valorar la validez y fiabilidad de los datos de eficacia y/o efectividad: utilización de listas guía para su evaluación. Curso de Farmacoeconomía on-line. 2001 Pharmacia