

Original

Guía para el análisis crítico de ensayos clínicos en Urología

Félix Millán Rodríguez

Urólogo. Máster en Metodología de la Investigación: Diseño y Estadística en Ciencias de la Salud. Fundación Puigvert. Barcelona. España

Resumen

El Ensayo Clínico Aleatorio (ECA) es el estudio con una mayor evidencia científica en Medicina. Aunque un urólogo puede no necesitar saber cómo se diseña un ECA, sí que le interesa saber cómo analizarlo de una forma crítica. En este trabajo, por un lado, se explican los elementos más importantes de un ECA: formulación de hipótesis, validez interna, relevancia de los resultados y validez externa, y por otro lado, se realiza una propuesta de cómo analizar un ECA a través de la pirámide de los "Niveles de información". Esta pirámide tiene dos características: primero, que para interpretarla se necesitan conocimientos clínicos más que estadísticos, y segundo, que a medida que se asciende por sus sucesivos niveles (significación estadística, resultado, intervalo de confianza, utilidad clínica y utilidad en la población) aumenta la información clínica. Asimismo, se comentan los métodos más frecuentemente empleados para magnificar los resultados hallados en un ECA: dar más importancia a la significación estadística que a la relevancia clínica, la forma gráfica de presentar los resultados, o el uso de variables poco relevantes desde el punto de vista clínico.

Palabras clave: Guía. Ensayo Clínico Aleatorio. Análisis. Evidencia Científica.

Guideline on critical analysis of urologist clinical trials

Abstract

The Randomized Clinical Trial (RCT) is the study offering the greatest scientific evidence in Medicine. Although urologists may not know how a RCT is designed, it is advisable for them to be able to analyze it critically. The present study on one hand explains the most important elements of a RCT: developing a hypothesis, internal validity, relevance of the results and external validity. On the other hand, a proposal is made as to how to analyze a RCT through the so-called "levels of information" pyramid. This pyramid has two features: (a) its interpretation requires clinical rather than statistical knowledge; and (b) as we move up its successive levels (statistical significance, results, confidence interval, clinical usefulness and usefulness in the population), the clinical information increases. Likewise, mention is made of the most common methods used to magnify the results obtained by a RCT: the assignment of greater importance to statistical significance than to clinical relevance, the graphical manner of presenting the results, or the use of variables of scant relevance from the clinical perspective.

Keywords: Guide. Randomized clinical trial. Analysis. Scientific evidence.

Existen diferentes tipos de estudios para evaluar la acción de una intervención o un tratamiento sobre una enfermedad. Los estudios experimentales, que se definen como estudios de cohortes en los que el investigador manipula la intervención y observa el efecto sobre un desenlace, constituyen el mejor método para evaluar dichos tratamientos o intervenciones. Dentro de ellos, el ensayo clínico aleatorio (ECA) representa el paradigma para estudiar intervenciones en Ciencias de la Salud. Ante esto hay que hacer dos consideraciones. Primero, que los trabajos a los que se debe prestar atención prioritaria sobre resultados de opciones terapéuticas en Medicina deben estar basados en ECAs. Y segundo,

que no es prudente aceptar los resultados de un ECA determinado sin haber realizado previamente una valoración crítica del mismo. Precisamente, el primer problema práctico que surge cuando se intentan aplicar los conocimientos estadísticos y metodológicos es valorar críticamente un ECA. Trasladado a un escenario real ¿estamos en condiciones de aportar una postura crítica frente a una publicación o una información de la industria farmacéutica que presente un ECA sobre Urología?

Habitualmente, lo primero que se piensa cuando se ha de abordar el análisis de un estudio con un tratamiento estadístico más o menos complejo es que nos hemos de enfrentar a una disciplina ardua.

Posiblemente, la causa está en la complejidad intrínseca de estos temas pero sobre todo en la forma de exponerlos. Habitualmente, los tratados de estadística o epidemiología enseñan cómo realizar un ECA pero, con menos frecuencia, cómo interpretarlo; y dado que la mayoría de urólogos no van a diseñar un ECA, pero sí que van a leer centenares de trabajos y comunicaciones a lo largo de su actividad, el objetivo del presente artículo es proporcionar al urólogo una guía práctica sobre cómo interpretar y evaluar un ECA en Urología.

EL ENSAYO CLÍNICO ALEATORIO

Un ECA compara intervenciones diferentes en *dos o más grupos* de sujetos. Como mínimo debe haber un grupo que recibe la intervención (grupo de estudio) y otro que recibe otra intervención o un placebo (grupo control). Por otro lado, un ECA es un *estudio experimental*. Esto quiere decir que es el investigador el que controla el factor de estudio, o sea que es él quien decide qué tratamiento dar a cada grupo, de qué forma o durante cuánto tiempo. Los estudios no experimentales se denominan observacionales porque el investigador se limita a observar la acción de unos factores de estudio que no controla. Aunque estos tienen una menor evidencia científica que los estudios experimentales, no siempre es posible diseñar un estudio experimental. Lo comprenderemos con el siguiente ejemplo. Si se quisiera valorar el efecto del tabaquismo en la aparición del tumor vesical, se debería hacer mediante un estudio observacional y no mediante un experimental, ya que éste supondría que el investigador asignaría el factor de estudio "tabaquismo" y habría de decidir el procedimiento para asignar quién debe fumar y quién no, lo cual es éticamente inviable. Por ello, la única solución sería realizar un estudio observacional, como por ejemplo, un estudio de cohortes prospectivo en el que se estudiaría una cohorte de fumadores y otra de no fumadores, y al cabo de un tiempo se contabilizaría la diferente incidencia de tumor vesical en ambos grupos. Finalmente, otro aspecto fundamental de un ECA es la *asignación aleatoria* de la intervención, con el objetivo de conseguir grupos comparables en todo excepto en la intervención asignada. Si esto se logra, cualquier diferencia en la variable respuesta se puede atribuir a la intervención estudiada.

Cuando queramos analizar un artículo sobre un ECA es recomendable seguir el esquema que iremos exponiendo en los sucesivos apartados (Tabla 1).

Tabla 1. Elementos a analizar en un ensayo clínico aleatorio

Formulación de la hipótesis

- ¿El estudio aporta algo nuevo?
- ¿El objetivo del estudio está bien definido?

Validez interna

- Asignación aleatoria de los sujetos.
- Seguimiento de los pacientes.
- Estrategia de análisis a tratar.
- Enmascaramiento.
- Comparabilidad inicial de los grupos

Relevancia de los resultados

- Magnitud del efecto
- Precisión
- Representación gráfica

Validez externa

- Diferencias entre los pacientes seleccionados y la población a tratar.
- Modificación del pronóstico.
- Relación entre beneficios, seguridad, coste.

Primero analizar la hipótesis que se quiere investigar, después la validez interna, en tercer lugar la utilidad clínica de los resultados y finalmente su validez externa. En este trabajo, los distintos ejemplos estarán basados en ECAs sobre el tratamiento médico de la HBP, ya que constituyen un buen modelo para aprender a interpretar estudios y porque por otro lado, es una de las patologías urológicas sobre la que más ECAs se han publicado.

Un primer vistazo: formulación de la hipótesis

Del mismo modo que cuando conocemos a alguien nuevo nos extiende su tarjeta de visita, o que en una librería un libro reclama nuestra atención a través del diseño de su portada, en un ECA o en cualquier otro estudio, el *título* es el primer elemento de contacto y por ello debería sugerir el objetivo y los resultados del trabajo. Si ha conseguido llamar nuestra atención, a continuación pasaremos a analizar la "Introducción". Es aquí donde debe exponerse claramente el *objetivo* del ECA y sus fundamentos. Dos son los errores más frecuentes: abrumar con una gran cantidad de citas bibliográficas perdiéndose en datos irrelevantes y no justificar o definir correctamente el objetivo del estudio.

La relevancia del estudio estará en función de si es un trabajo que aporta algo nuevo o, por si lo contrario, se limita a repetir estudios ya realizados en el pasado; en este último caso, el interés será menor a no ser que los resultados sean diferentes a

lo publicado previamente. Por ejemplo, actualmente ya no tiene mucho interés realizar un ECA para valorar la tolerancia de los alfabloqueantes o de los inhibidores de la 5 alfa-reductasa. En cambio, sí que puede ser novedoso estudiar la eficacia de estos fármacos en indicaciones nuevas. En cualquier caso, la existencia de diferentes ECA que tengan el mismo objetivo puede servir ulteriormente para la realización de un meta-análisis que valore conjuntamente todos los estudios publicados previamente.

Por otro lado, hay que recordar que un ECA tiene que tener claramente definido el objetivo del estudio y la hipótesis a contrastar. En este sentido, es incorrecto no especificar la hipótesis del estudio o hacerlo de forma genérica o inespecífica. De este modo, es bastante habitual encontrarnos con objetivos como el siguiente ejemplo:

[a] "El objetivo del presente estudio es valorar la eficacia del fármaco A en el tratamiento de la hiperplasia prostática benigna".

Éste sería un ejemplo de un objetivo enunciado de forma vaga, ya que se nos plantean muchos interrogantes. ¿Cómo se valora la eficacia? ¿Se puede administrar cualquier dosis de fármaco A? ¿Qué casos son tratados: próstatas grandes, asintomáticas, complicadas...?

Otro error es pretender analizar numerosos objetivos en un mismo estudio. Se demuestra estadísticamente que a medida que aumenta el número de variables a estudiar, aumenta también la probabilidad de encontrar resultados estadísticamente significativos debido al azar, riesgo que se incrementa si el número de pacientes estudiados es bajo. Éste podría ser un ejemplo:

[b] "Los objetivos de nuestro estudio fueron evaluar la eficacia del fármaco A en el tratamiento de la HBP, su acción sobre la calidad de vida, su posible influencia en la vida sexual del paciente, así como su coste-beneficio".

A veces se tiende a pensar que cuantos más objetivos estudie un trabajo mayor calidad tiene, mientras que lo único que hay que pedirle es que la hipótesis que se quiere demostrar esté bien especificada y que se desarrolle de una forma metodológicamente correcta. Un objetivo enunciado correctamente podría ser el del ejemplo siguiente:

[c] "El objetivo de este estudio fue valorar el efecto de la administración durante 12 meses de 5 mg diarios del fármaco A vía oral, sobre el flujo máximo urinario en pacientes de 50 a 75 años afectos de hiper-

plasia benigna prostática sin patología asociada, que presentaban una sintomatología leve moderada definida como un valor de IPSS de 5 a 15 puntos".

En este ejemplo vemos que sólo hay una hipótesis de estudio (eficacia del fármaco) y que está perfectamente definida la dosificación, la población del estudio y la forma de evaluación. Si tras realizar el estudio se observa que el fármaco A tiene una eficacia mayor que el placebo, podemos concluir que es eficaz para el tratamiento de la HBP en pacientes con esa edad y esas características. Ahora bien, en un estudio como el del enunciado [a], si la diferencia es también estadísticamente significativa, ¿a qué edad o tipos de pacientes o de sintomatología podremos inferir que el tratamiento con el fármaco A es eficaz?. Después de estos ejemplos, seremos conscientes de la importancia que tiene que cualquier tipo de estudio debe tener un objetivo bien fundamentado y claramente expuesto. Sin embargo, aún quedan por analizar más elementos que exponemos a continuación.

Validez interna: ¿El efecto observado se debe al tratamiento y no a otras variables?

En un ECA siempre existen dos hipótesis. La hipótesis nula (H0) es la que se pretende rechazar y la hipótesis alternativa (H1) la que se pretende demostrar. Pongamos un ejemplo:

[d] "H0: Los pacientes afectos de HBP tratados mediante el fármaco B no presentan un aumento del flujo máximo urinario respecto a los tratados con placebo".

[e] "H1: Los pacientes afectos de HBP tratados mediante el fármaco B presentan un aumento del flujo máximo urinario respecto a los tratados con placebo".

Supongamos que ya se ha realizado el ECA, y con los datos obtenidos tras aplicar unas pruebas estadísticas determinadas, obtenemos unos resultados que indican que la diferencia es estadísticamente significativa por lo que aceptamos la H1. En este ejemplo, el fármaco B es más eficaz que el placebo en el tratamiento de la HBP. No obstante, lo primero que hay que plantearse es si no podemos haber llegado a una conclusión falsa, o sea que la hipótesis cierta sea la H0 y que por error se haya aceptado la H1. Este es el llamado error alfa (probabilidad de aceptar la H1 cuando la verdadera es la H0) (Fig. 1). Habitualmente se acepta un error de un 5%; por eso siempre se dice que aceptamos la H1

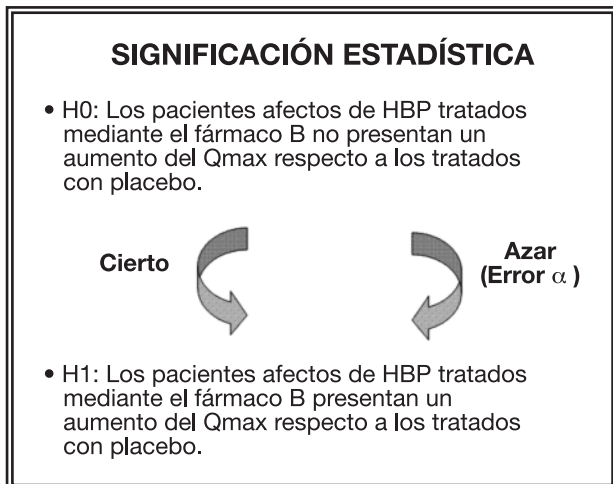


FIGURA 1. Concepto de significación estadística y error alfa.

cuando la p es menor de 0,05. Esto significa que la probabilidad de que los resultados observados hayan sido debidos al azar es menor del 5%.

Llegados a este punto, siempre se ha de comprobar si el diseño del estudio está realizado correctamente, y por lo tanto, el efecto observado es debido al tratamiento estudiado y no a otras variables. Esto es lo que se llama el estudio de la *validez interna* de un ECA: evaluar si los resultados son válidos y no han sido debidos a una estimación sesgada del efecto del tratamiento. Para valorarlo han de analizarse los siguientes puntos:

1. Asignación aleatoria de los sujetos

Supongamos que un ECA que compara la eficacia de un fármaco C versus un fármaco D en el tratamiento de la HBP, demuestra que los pacientes tratados con D tienen una mejoría sintomatológica estadísticamente superior a los tratados con C. A primera vista, lo correcto sería concluir que D es más eficaz que C en el tratamiento de la HBP. No obstante, imaginemos que al analizar los dos grupos se observa que los pacientes tenían un tamaño prostático medio de 47 y 28 gramos en el grupo del fármaco C y D respectivamente. En este caso, se plantearían dudas razonables sobre si la diferencia observada es debida realmente a una mayor eficacia del fármaco D o a que como los pacientes tratados con el fármaco C tenían unas próstatas mayores, no ha sido posible demostrar una mayor eficacia del fármaco C. Cualquier urólogo se daría cuenta rápidamente de la escasa validez del estudio porque sabe que el tamaño prostático influye en el resultado.

¿Pero qué pasaría con otras posibles variables desconocidas que puedan influir en el resultado y que no podamos evitar?

Por todo lo expuesto, se comprenderá que lo ideal en un ECA es que los grupos a comparar estén equilibrados en todas las posibles variables influyentes en el resultado (conocidas o no), y se parezcan en todo menos en la variable tratamiento asignado. De esta forma se podrá aceptar que la diferencia en el resultado si la hubiere, sería debida exclusivamente al tratamiento asignado.

En esto radica la importancia de la asignación de los sujetos al azar, ya que es la única forma de poder asegurar una distribución por igual en los grupos de estudio de todas las posibles variables pronósticas conocidas o no. Al ser un proceso aleatorio, ni el investigador, ni el paciente, ni nadie más influirá en la asignación de los grupos de tratamiento. Si no se realiza la asignación de forma aleatoria se comete un *sesgo de comparabilidad*.

Por otro lado, es importante no olvidar dos cuestiones. La primera, que la asignación aleatoria es la que proporciona una mayor garantía de distribución equitativa pero no la garantiza siempre. Por ello, en todo ECA es importante realizar siempre una comparación de la distribución de las posibles variables pronósticas en los grupos de tratamiento, ya que el azar puede causar a veces una mala jugada. En segundo lugar, la asignación aleatoria ha de realizarse siempre después de que el paciente haya cumplido los criterios de inclusión y haya firmado el consentimiento. Cualquier estudio que no lo realice de esta forma incurrirá en un grave problema de validez del mismo.

2. Seguimiento de los pacientes

Como se acaba de ver, si se ha realizado una asignación aleatoria se garantiza la comparabilidad de los grupos. Sin embargo, si durante el curso del estudio hay pacientes que por cualquier motivo abandonan y dejan el estudio, esta comparabilidad inicial puede peligrar, por lo que puede haber dudas sobre si el resultado final observado es debido al tratamiento o a la presencia de características diferentes en los grupos estudiados. Imaginemos que en el anterior estudio abandonan igual número de pacientes de ambos grupos. Si los pacientes que abandonan son iguales en todas las características quizás no se vea artefactuado el estudio. Pero, ¿qué ocurriría si, por ejemplo, los pacientes tratados con

el fármaco C que abandonan tienen próstatas de un tamaño significativamente mayor? En el grupo del fármaco C quedarían pacientes con próstatas de menor tamaño y si al final del estudio este grupo presenta una mejoría de los síntomas significativa, no sabríamos si atribuirla al efecto del fármaco C o a la falta de comparabilidad por la pérdida de estos pacientes. Con esto vemos que es tan importante conocer el número de los pacientes que abandonan como las características de los mismos.

El sesgo que se produce es mayor cuanto más diferentes son las características de los sujetos que abandonan, en comparación a los que completan el estudio. Este hecho es mucho más grave si el motivo por el que abandonan el estudio está relacionado con la intervención. Por otro lado, una cuestión que surge a menudo, es determinar el número máximo de pacientes que pueden abandonar un estudio para considerarlo válido. Es difícil establecer un criterio único ya que como se ha explicado, puede ser más determinante el motivo por el que abandonan que el número de sujetos que lo hacen. Además, esto dependerá del tipo de estudio. Supongamos por ejemplo que se han diseñado dos ECA: en uno se quiere encontrar resultados significativos con una diferencia del 5% y en el otro del 20%. Si en ambos estudios hay un 10% de abandonos, el que puede presentar mayores problemas de validez es el primer estudio ya que el porcentaje de abandonos es el doble de la diferencia que queremos hallar. En cambio en el segundo estudio, este mismo porcentaje de abandonos puede ser más tolerable. En cualquier caso, un ECA debe *informar siempre* del número de pacientes que han abandonado, y de las características y la respuesta de éstos a la intervención.

3. Estrategia de análisis

Si el hecho de que haya pacientes que abandonen, aporta problemas al análisis, ¿qué ocurre cuando, por el motivo que sea, hay pacientes que han recibido la intervención que no les correspondía?. Lo analizaremos con el siguiente ejemplo (Tabla 2). Un ECA pretende comparar en 400 pacientes la eficacia entre el fármaco E y el fármaco F, de forma que se asigna aleatoriamente 197 al grupo de E y 203 al de F. Sin embargo 12 pacientes a los que se les asignó el fármaco E acabaron tomando el F y 17 a la inversa. El análisis de estos resultados se puede llevar a cabo de 3 maneras diferentes:

Tabla 2. Estrategia de análisis

Estudio comparativo entre el fármaco E y el fármaco F con 400 pacientes.

Se asignan 197 a E
Se asignan 203 a F

De los 197 asignados a E, toman correctamente E 185
De los 203 asignados a F, toman correctamente F 186
De los 197 asignados a E, toman erróneamente F 12
De los 203 asignados a F, toman erróneamente E 17

Análisis según eficacia biológica. Se comparan los 185 que tomaron correctamente E versus los 186 que tomaron correctamente F.

Análisis según tratamiento. Se comparan los 202 (185 + 17) que acabaron tomando E versus los 198 (186 + 12) que acabaron tomando F.

Análisis según intención de tratar. Se comparan los 197 asignados al grupo de E versus los 203 asignados al grupo de F.

– *Eficacia biológica:* Se trata de analizar y comparar sólo los que realmente siguieron el tratamiento previamente asignado. En este caso se compararían los 185 pacientes (197–12) que se les asignó el fármaco E, y realmente la tomaron frente a los 186 (203–17) que se les asignó y tomaron el fármaco F.

– *Según tratamiento:* Se comparan los casos que tomaron un determinado fármaco independientemente del que se les hubiera asignado previamente. En nuestro ejemplo se analizarían los 202 (185 + 17) que tomaron el fármaco E, frente a los 198 (186 + 12) que acabaron tomando el fármaco F.

– *Según intención de tratar:* Se analizan los casos según la asignación previa establecida independientemente del tratamiento que hayan tomado después. En este estudio se compararían los 197 a los que se les asignó el fármaco E, frente a los 203 que se les había asignado el fármaco F.

Aunque a priori parezca lo menos razonable, la única estrategia adecuada de análisis es el *análisis por intención de tratar*. El motivo es porque es el único método que respeta la comparabilidad inicial de los grupos, conseguida mediante la asignación aleatoria. Cualquiera de las otras estrategias conlleva una nueva reagrupación que puede desequilibrar la comparabilidad de los nuevos grupos resultantes. Por todo ello, cualquier ECA que no utilice un análisis por intención de tratar estará sesgado por lo que su validez será discutible.

4. Enmascaramiento

Aunque la mayoría de las variables que se analizan suelen ser objetivas, el que el investigador o el paciente conozca a qué grupo de tratamiento pertenece puede tener una influencia en la evaluación de la respuesta. Para evitar esto se llevan a cabo las técnicas de enmascaramiento. La más sencilla es el "simple ciego" en la que el paciente desconoce el tratamiento que recibe pero no el investigador. Cuando el investigador desconoce también a qué grupo es asignado cada paciente se denomina "doble ciego". Si además la persona que analiza los datos también lo desconoce estamos ante un estudio a "triple ciego". Hay estudios en los que el enmascaramiento es más fácil de realizar que en otros; así en aquellos en los que la intervención es un fármaco, se administran los diferentes fármacos o los placebos en la misma presentación y dosificación. Sin embargo, si lo que se quiere diseñar es un ECA que compare una nueva técnica quirúrgica en HBP versus a la RTU de próstata, es muy difícil realizar un enmascaramiento ya que tanto el investigador como el paciente han de conocer qué técnica quirúrgica se ha aplicado.

5. Comparabilidad de los grupos al iniciar el estudio

Cuando se ha hablado de la asignación aleatoria se ha comentado la importancia de que los grupos de estudio sean comparables en todas las variables, excepto en la intervención. No obstante, la asignación aleatoria no asegura de forma fehaciente que los grupos acaben siendo equiparables, por lo que siempre es recomendable que un ECA informe y compare las características de los grupos al inicio del estudio.

Relevancia de los resultados

En el apartado anterior se ha analizado cómo saber si un ECA es válido, es decir, si los resultados obtenidos se pueden atribuir de una forma razonable únicamente a la acción de las intervenciones realizadas y no a otras variables. Se trata de un análisis meramente metodológico pero a un clínico lo que más le interesa es saber si los resultados del estudio le van a aportar alguna mejora desde el punto de vista clínico. Dicho de otro modo, el que decidamos utilizar un tratamiento o una intervención porque es más eficaz que otro, no ha de estar fundamentado únicamente en una prueba estadística de significación sino que ha de basarse también en la relevancia clínica de los resultados encontrados.

¿Qué es exactamente la significación estadística?

Como se comentaba en el capítulo de la validez interna, se trata de la probabilidad de que los resultados obtenidos hayan sido debidos al azar. Por consenso se acepta en la comunidad científica que es razonable aceptar los resultados de un estudio cuando la probabilidad de que sean debidos al azar sea inferior al 5%, de ahí que se hable de una $p < 0,05$. De todas maneras, se trata de un resultado estadístico que no tiene expresión en términos clínicos. Así, por ejemplo, el siguiente resultado que es correcto estadísticamente aporta muy poca información clínica:

[f] "Los resultados del presente estudio demuestran que el fármaco G consigue un aumento del flujo máximo estadísticamente significativo ($p=0,001$), con respecto al placebo".

Sin embargo, lo que va a tener utilidad clínica es la relevancia de los resultados que se evalúan analizando la magnitud del efecto y la precisión.

1. Magnitud del efecto

A menudo, el criterio para evaluar el resultado de un estudio es fijarse en la significación estadística, es decir, en el valor de la "p". Como acabamos de ver, valorar únicamente esto no es lo más correcto porque la "p" nos informa de aspectos estadísticos y no clínicos. Por otro lado, el hecho de tomar como criterio de análisis el valor de la "p" puede conllevar otros problemas. Supongamos, por ejemplo, que se quiere comparar los resultados de dos ECA (Tabla 3). Ambos alfabloqueantes aumentan significativamente el flujo máximo respecto al placebo; el alfabloqueante X con una $p=0,04$ y el alfabloqueante Y con una $p=0,00001$. Una actitud frecuente es considerar que el alfabloqueante Y es mucho mejor que el X porque "su p" es mucho más pequeña. ¡Esto es un error muy grave! De nuevo es preciso hacer énfasis en que la significación estadística es un concepto estadístico y no clínico, y lo único que nos dicen estos datos es que la probabilidad de que estos resultados sean debidos al azar

Table 3. Ejemplo sobre significación estadística y magnitud del efecto

Alfa-bloqueante	Significación estadística	Aumento medio del flujo máximo (ml/s)	Intervalo de confianza 95% (ml/s)
X	$p = 0,04$	3,3	2,7 - 3,9
Y	$p = 0,00001$	1,5	0,8 - 2,2

es del 4% en el estudio del alfabloqueante X, y de 1/100000 en el otro estudio. Sin embargo, esto no proporciona información clínica sobre la magnitud del efecto. Para conocer el efecto ha de analizarse cuánto aumenta el flujo máximo en ambos estudios y con qué precisión. En este caso, se observa que el alfabloqueante X causa un aumento medio del flujo máximo de 3,3 ml/s, mientras que con el alfabloqueante Y el aumento medio del flujo máximo es menor 1,5 ml/s. Con este ejemplo se comprende cómo un estudio que aparentemente tiene “una p muy buena” no tiene por qué llevar asociada una relevancia clínica importante.

2. Precisión

Supongamos que en el estudio previo del fármaco G versus placebo se demuestra que hay un aumento del flujo máximo estadísticamente significativo ($p=0,001$) de 2,1 ml/s. Se expresaría de esta manera:

[g] “Los resultados del presente estudio demuestran que el fármaco G consigue un aumento del flujo máximo de 2,1 ml/s estadísticamente significativo ($p=0,001$), con respecto al placebo”.

Este trabajo ha sido basado en una muestra concreta de la población estudiada. Si se realizara otro estudio idéntico con una muestra diferente de la misma población también saldría otro resultado estadísticamente significativo, pero el incremento del flujo máximo no tendría por qué ser exactamente de 2,1 ml/s sino que podría ser de 1,9 ó 2,3 ml/s por ejemplo. Sucesivamente, el resultado volvería a ser diferente con distintas muestras que se usasen de la misma población, pero siempre se movería en un rango de valores. De aquí viene el concepto de intervalo de confianza del 95% (Fig. 2). Éste sería el rango de valores entre los que fluctuaría el resultado de una prueba en el 95% de los casos en que se aplicara la misma. Por ejemplo, si en este estudio el intervalo fuera 1,9-2,3 ml/s el resultado se expresaría de la siguiente manera:

[h] “Los resultados del presente estudio demuestran que el fármaco G consigue en el 95% de los casos un aumento del flujo máximo de 1,9 a 2,3 ml/s, con respecto al placebo”.

Otro aspecto importante, es que cuanto mayor sea la muestra utilizada en un ECA menor será el intervalo de confianza, es decir, será más preciso. Así, en nuestro ejemplo, si la muestra hubiera sido más pequeña en vez de obtener el resultado 2,1

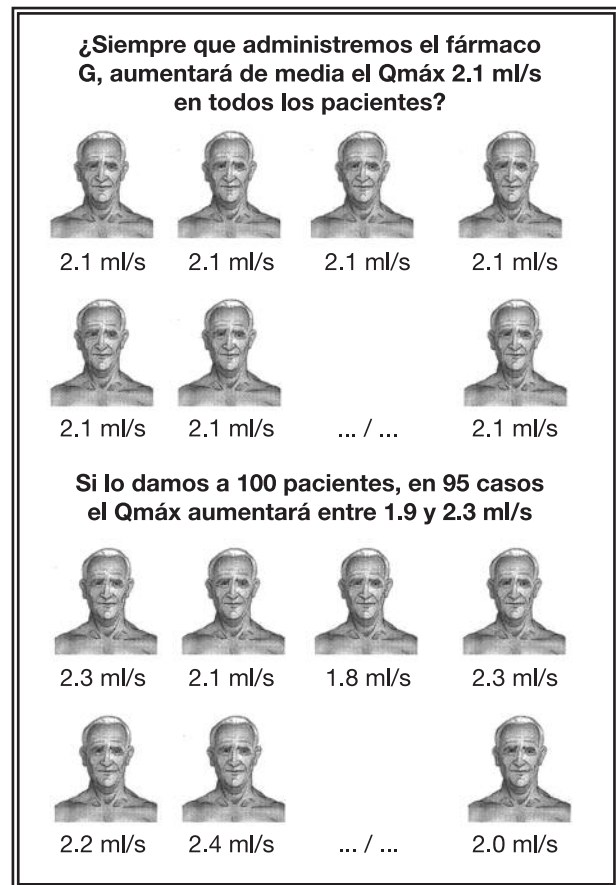


FIGURA 2. Concepto de intervalo de confianza del 95%.

ml/s (IC95%=1,9-2,3) se habría obtenido, por ejemplo, 2,1 ml/s (IC95% = 1,4-2,8), con lo que el rango dentro del que oscilan los valores sería más ancho y por lo tanto menos preciso.

Por todo esto, cuando se analiza un ECA es muy importante comprobar que el resultado, no sólo esté expresado en términos de significación estadística, sino también en términos clínicos y concretamente mediante el intervalo de confianza del 95%. Analizando este intervalo se podrá conocer la variación mínima y máxima a esperar tras aplicar el tratamiento, y también podremos valorar la precisión viendo si el rango es muy estrecho o muy ancho. Lógicamente, cuando más estrecho sea el rango, será mucho mejor ya que será un resultado más preciso.

Si nos fijamos en las diferentes formas de expresar el mismo resultado [f], [g] y [h], han ido incorporando cada vez más información clínica. Así la [f] únicamente comenta que el resultado es estadísticamente significativo pero no dice nada de la utilidad clínica del resultado. En cambio, con la defini-

ción [h] no sólo sabemos que es estadísticamente significativo sino que también informa de que usando el fármaco G, se consigue en el 95% de los casos un aumento del flujo máximo como mínimo de 1,9 ml/s. Ésta es la ventaja que aporta el presentar los resultados con intervalos de confianza, el dar una información útil desde el punto de vista clínico, ya que lo que interesa al urólogo, no es sólo si el resultado es significativo desde el punto de vista estadístico, sino si va a ser útil desde el punto de vista de la práctica diaria.

Llegados a este punto, hemos ido subiendo tres peldaños de información: el valor "p", el valor del resultado y el valor del intervalo de confianza. Cada nivel nos aporta más información clínica que el previo. Sin embargo, aún se puede subir un peldaño más que es el de la utilidad clínica. De hecho, es lo que tiene que interesar a un urólogo a la hora de decidir prescribir o no un nuevo tratamiento en HBP. Por ejemplo, cuando se valora la variable "flujo máximo" la mayoría de los ECA sobre medicación en HBP nos ofrecen resultados estadísticamente significativos, con sus respectivos intervalos de confianza. Sin embargo, los aumentos de flujo máximo obtenidos con el tratamiento suelen oscilar entre 1,5 ó 3 ml/s. Ante este hecho, los urólogos debemos plantearnos ¿el hecho de que un paciente aumente su flujo máximo en 1.5 o incluso 3 ml/s lo va a notar en su forma de orinar? Es más, ¿esta diferencia de 1,5 ó 3 ml/s es clínicamente relevante cuando está demostrado que si se realizan dos flujometrías consecutivas a un mismo paciente pueden presentar también esta misma variabilidad de 2 ml/s? Con esto se pretende llamar la atención sobre el hecho de que un fármaco puede conseguir unos resultados estadísticamente fabulosos, pero que sean clínicamente poco relevantes. Para analizar la relevancia clínica de los resultados, el urólogo no necesita de conocimientos estadísticos sino de sentido común y experiencia clínica, por lo que no ha de sentir temor a la hora de analizar críticamente los resultados de un ECA. Se ha comentado el ejemplo del "flujo máximo" porque en los ECA en HBP es quizás uno de los más evidentes en cuanto a discrepancia entre relevancia estadística y relevancia clínica. Sin embargo, existen muchas otras variables que se han usado para evaluar la acción de un fármaco en HBP: mejoría del flujo máximo, mejoría de puntuación de la escala IPSS, reducción de tamaño prostático, reducción de la incidencia de retenciones

urinarias, reducción de las indicaciones de cirugía... En cualquier caso, independientemente de la variable estudiada siempre se ha de valorar la implicación clínica de los resultados obtenidos.

3. Representación gráfica de los resultados

Acabamos de estudiar que la "p" sólo nos ha de servir para comprobar que un resultado es estadísticamente significativo; una vez que lo es, para valorar el efecto clínico no hemos de fijarnos más en el valor de la "p", sino en el resultado de la prueba y su intervalo de confianza. Por ello, a continuación, ha de valorarse la repercusión de estos resultados aplicados a la práctica clínica diaria. Sin embargo, se ha de ser cauteloso con aquellos trabajos que únicamente remarcan que son estadísticamente significativos pero que no informan, o lo hacen de pasada, sobre el resultado de la prueba, su precisión o su relevancia clínica.

Por otro lado, existen otros trabajos que para enmascarar la baja relevancia clínica de la eficacia de un fármaco, lo que hacen, además de focalizar la atención en la "p", es magnificar de forma visual el efecto de dicho fármaco en la gráfica de presentación de resultados. Sin ir más lejos, siguiendo el ejemplo del tratamiento médico de la HBP, tenemos dos trabajos clásicos de Lepor y Roehrborn, sobre el efecto de la tamsulosina y terazosina, respectivamente. La Figura 3A, muestra la gráfica que se publicó en su día del efecto de la tamsulosina y el placebo en el aumento del flujo máximo. A este trabajo, y a otros muchos sobre el tratamiento médico de la HBP, cabría realizarle muchas objeciones; por ejemplo, ¿es correcto analizar en HBP resultados a sólo 18 semanas de seguimiento? Sin embargo, nos centraremos en el tema de la gráfica de representación de resultados. En la figura se observa un aumento medio del flujo máximo de 1,5-2 ml/s y de 0,25-0,75 ml/s en los grupos tratados con tamsulosina y placebo, respectivamente. El artículo en cuestión, nos informa de que esta diferencia media de flujo máximo inferior a 2 ml/s es estadísticamente significativa. Obviamente, esta diferencia, por muy estadísticamente significativa que sea, no tiene relevancia clínica, por lo que una forma de magnificarla es presentando los resultados en una gráfica que tabula el flujo máximo desde 0 a 2,5 ml/s. Dado que en Urología, aceptamos el valor de 15 ml/s como límite de la normalidad, ¿qué pasaría si estos mismos resultados se hubieran representado en

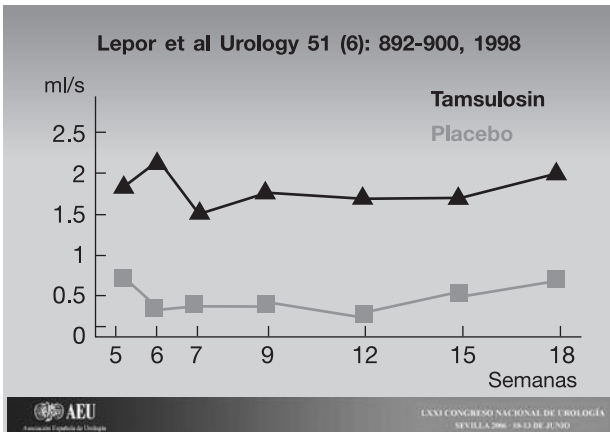


FIGURA 3A. Adaptación de la Figura 1B de Lepor y col; Urology 51(6):892-900, 1998.

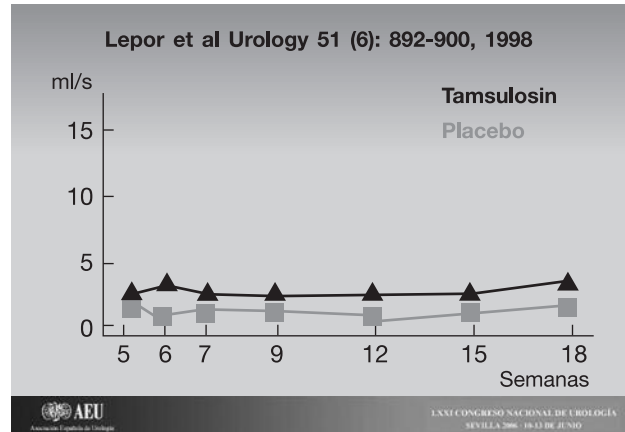


FIGURA 3B. Rediseño de la Figura 3A tras escalar el flujo máximo de 0 a 15 ml/s.

una gráfica tabulada de 0 a 15 ml/s? Esto es lo que hemos realizado en la Figura 3B. Al observarla, se comprueba que el efecto visual de la eficacia de la tamsulosina respecto al placebo, sin haber cambiado los valores, es mucho más modesto de lo que se nos presentó en la figura que se publicó en el artículo (Fig. 3A). Este mismo truco visual se empleó también en el artículo que comparaba la eficacia de la terazosina y el placebo (Fig. 4A), donde la gráfica se tabuló con unos valores de flujo máximo de 0 a 1.4 ml/s. De esta forma, si estos valores se vuelven a representar en una gráfica tabulada de 0 a 15 ml/s (Fig. 4B), aunque las diferencias son estadísticamente significativas, las líneas de la terazosina y del placebo hasta casi se superponen visualmente. Con estas estrategias, que son bastante habituales, no se está mintiendo sobre los resultados, pero sí presentándolos de una forma visual que nos incita

a interpretar que tienen una relevancia clínica superior a la que realmente tienen.

Validez externa: Utilidad práctica de los resultados

Una vez que un ECA ha demostrado tener validez interna y que sus resultados son clínicamente relevantes, ha de evaluarse si el tratamiento o la intervención estudiada puede ser de utilidad para manejar nuestros propios pacientes. Nos ayudará a ello el análisis de los siguientes elementos.

1. Diferencias entre los pacientes seleccionados y la población a tratar

La muestra es un conjunto de pacientes representativos de la población que se pretende estudiar. Esta muestra viene determinada por los criterios de inclusión. Cuanto más restrictivos sean los criterios

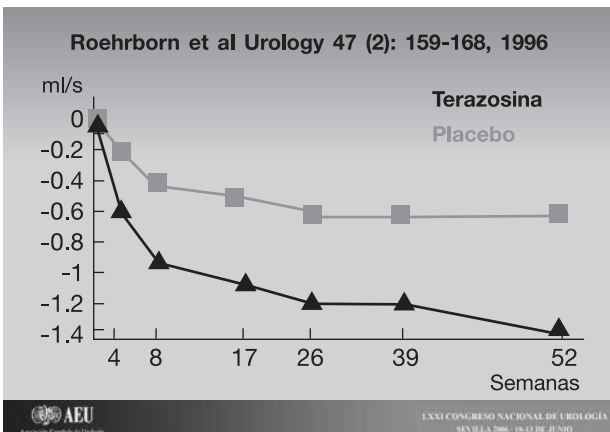


FIGURA 4A. Adaptación de la Figura 2D de Roehrborn y col; Urology 47(2):159-168, 1996.

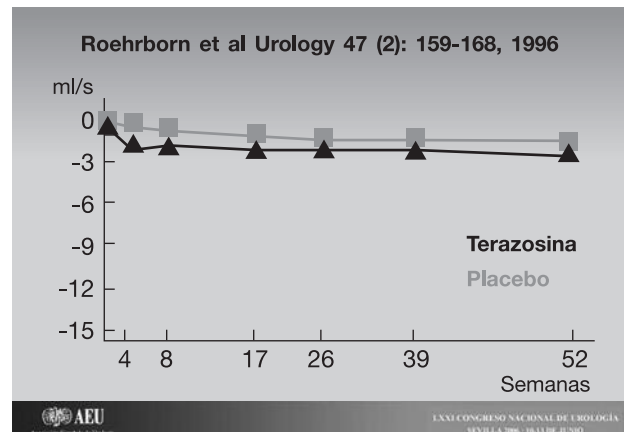


FIGURA 4B. Rediseño de la Figura 4A tras escalar el flujo máximo de 0 a 15 ml/s.

de inclusión la muestra será más homogénea y el estudio ganará en validez interna. Por el contrario, cuanto más laxos sean los criterios de inclusión la muestra será más heterogénea y perderá validez interna a costa de ganar validez externa (capacidad de aplicar los resultados a la población general). No obstante, lo que interesa en un ECA es que el estudio tenga una buena validez interna para que realmente pueda evaluar aquello que pretende evaluar. Por ello, en el capítulo de la formulación de hipótesis se comentó que un ECA no puede estar basado en hipótesis vagas [a], sino en hipótesis concretas y bien especificadas [c]:

[a] "El objetivo del presente estudio es valorar la eficacia del fármaco A en el tratamiento de la hiperplasia prostática benigna".

[c] "El objetivo de este estudio fue valorar el efecto de la administración, durante 3 meses de 5 mg diarios del fármaco A vía oral sobre el flujo máximo urinario, en pacientes de 50 a 75 años afectados de hiperplasia benigna prostática sin patología asociada, que presentaban una sintomatología leve moderada definida como un valor de IPSS de 5 a 15 puntos".

En consecuencia, el resultado de un estudio sólo puede ser aplicable a una población que tenga las características de los sujetos estudiados. Así, si el ECA con el objetivo [c] demuestra la eficacia del fármaco A, no se puede generalizar el resultado y concluir que el fármaco A aumenta el flujo máximo de todos los pacientes afectados de HBP. Lo correcto es decir que aumenta el flujo urinario máximo en los pacientes de 50 a 75 años afectados de HBP, sin patología asociada que presentan una sintomatología leve-moderada definida como un valor de IPSS de 5 a 15 puntos. La eficacia del fármaco A no se puede extrapolar a pacientes de otras edades, con patologías asociadas, o con obstrucción severa porque no han sido incluidos en este estudio. El motivo es porque existe la duda razonable de que si se hubieran incluido pacientes de estas características el resultado podría haber sido estadísticamente diferente.

¿Cuál es el problema que nos encontramos habitualmente con los ECAs en general, y con los ECAs de HBP en particular, a la hora de extrapolar los resultados? Principalmente, que las características de los pacientes que suelen participar en los ECA son muy diferentes de nuestros pacientes de la práctica clínica diaria. Habitualmente se incluyen pacientes entre 50 y 70 años, con próstatas de pequeño tamaño, sin infección asociada, sin diabe-

tes, sin enfermedades neurológicas, con valores de PSA bajos, con obstrucción leve... Mientras que en la realidad diaria encontramos muchos pacientes ASA 4, diabéticos, con antecedentes de accidente vascular cerebral, que toman dicumarínicos, algunos con cáncer de próstata asociado, la mayoría por encima de 70 años, con próstatas de todos los tamaños y muchas veces con obstrucción importante. En conclusión, un ECA siempre ha de informar de las principales características de los pacientes incluidos en el estudio para que el clínico pueda razonar sobre el grado de aplicabilidad de los resultados en función de la similitud o no con los pacientes que trata habitualmente.

2. Modificación del pronóstico

Con frecuencia, un ECA se limita a valorar únicamente las variables del estudio pero no hemos de olvidar que existen otras variables o desenlaces que pueden acabar teniendo importancia en el pronóstico del paciente. De esta forma es de interés conocer si la intervención estudiada modificará el pronóstico o el curso de la enfermedad. En primer lugar, muchos ECA en HBP están diseñados de forma que la eficacia se valora a los 3 meses, o como mucho a los 6 meses. En estos casos, cabría preguntarse si este efecto beneficioso observado inicialmente se va a mantener a largo plazo o si por el contrario va a ir desapareciendo. Por esto, ante una entidad como la HBP que tiene una evolución lenta y larga en el tiempo, ¿puede ser evaluado correctamente el efecto de un fármaco en la HBP con estudios con seguimiento inferiores a un año?

En segundo lugar, el efecto beneficioso de una intervención sobre la HBP aparte de mejorar unas variables objetivas (por ejemplo: flujo máximo, puntuación del cuestionario IPSS) ¿va a modificar la historia natural de la enfermedad? Imaginemos que un ECA compara la eficacia de un fármaco Z frente a placebo, y que a los 3 meses el fármaco Z consigue un incremento del flujo máximo de 3,4 ml/s (2,5-4,3), y el placebo de 1,5 ml/s (0,8-2,2) ($p=0,02$). En este caso el fármaco Z incrementa de forma significativa el flujo máximo respecto del placebo. No obstante, supongamos que al cabo de un año el 15% de los pacientes tratados con Z y el 21% de los tratados con placebo caen en retención, y han de ser intervenidos quirúrgicamente. Tendría que analizarse si esta diferencia es estadísticamente significativa. Si no lo fuera, nos encontraríamos ante un fár-

maco que es eficaz para incrementar el flujo máximo pero que no influye en la historia natural de la HBP, ni en el pronóstico de la misma porque los pacientes acaban siendo igualmente intervenidos independientemente del tratamiento asignado. Por todo ello, a la hora de analizar un ECA en HBP, es de suma importancia valorar el *tiempo de seguimiento* y si el tratamiento *modifica la historia natural de la HBP*.

3. Relación entre beneficios, seguridad y coste

Para valorar la aplicabilidad de un tratamiento a nuestros propios pacientes, no sólo han de evaluarse los beneficios, sino también la seguridad del fármaco y sus posibles efectos adversos. Por otro lado, un aspecto que cada día se va a tener más presente es el coste económico. Si existe un fármaco que tiene una eficacia similar que otro pero que es más caro que éste, lo más recomendable será prescribir el más barato a igualdad de eficacia. Finalmente, introduciremos un término muy importante y que tendría que ir apareciendo en los ECAs de calidad. Se trata del concepto de *número de pacientes a tratar (NPT)*. Se define como el número de pacientes que ha de recibir un tratamiento para conseguir cierto resultado. Lo entenderemos siguiendo el anterior ejemplo del ECA del fármaco Z versus placebo. Supongamos que la diferencia entre el 15% y el 21% de intervenciones al cabo de un año de tratamiento es estadísticamente significativa. Ante este resultado, se podría concluir que los pacientes tratados con el fármaco Z presentan una reducción significativa de la indicación quirúrgica por retención al cabo de un año, en comparación a los que han recibido placebo. En este caso, nos podríamos plantear ¿cuántos pacientes he de tratar con Z para evitar una cirugía por retención al cabo de un año? Este dato nos lo proporciona el NPT. En este ejemplo, no sería lo mismo un NPT de 20 que de 300. Un NPT = 20 significaría que por cada 20 pacientes que tratáramos con el fármaco Z, se evitaría una intervención por retención al cabo de un año de tratamiento. Se podría calcular el coste de la medicación y el de la cirugía, pero a priori parecería razonable. En cambio, si para ahorrar esa misma intervención por retención han de tratarse 300 pacientes con Z, quizás no estuviera justificado el coste económico y los posibles efectos adversos. Por otro lado, a la hora de valorar el NPT no sólo ha de tenerse en cuenta el valor en sí, sino también el concepto en que se basa.

Así, cuando se valora la NPT para evitar una progresión de la enfermedad en HBP, no es lo mismo definir progresión como un empeoramiento de la sintomatología en el score I-PSS, que un aumento del volumen prostático, una retención urinaria o la necesidad de realizar cirugía.

PROPUESTA DE INTERPRETACIÓN DE UN ENSAYO CLÍNICO (NIVELES DE INFORMACIÓN)

Hasta aquí se han ido desarrollando a través de múltiples ejemplos, las definiciones y los fundamentos teóricos de los principales aspectos a valorar en el análisis de un ECA, que se hayan resumidos en la Tabla 1. No obstante, dado que a muchos urólogos les puede resultar difícil un análisis profundo de un ECA, sobre todo los aspectos referidos a la validez interna, se quiere acabar el presente trabajo haciendo una propuesta de guía para analizar un ECA basado en la repercusión clínica de sus resultados más que en una valoración estadística. De este modo, esta propuesta de análisis puede ser de utilidad para cualquier urólogo sin conocimientos estadísticos que quiera analizar un ECA.

Para exponerla, nos basaremos en un hipotético ECA que compara la eficacia del fármaco X respecto al placebo, en el incremento del flujo máximo en HBP, cuyos resultados más importantes se hayan expuestos en la Tabla 4. Se propone el análisis siguiendo la pirámide de los niveles de información (Fig. 5), que tiene dos características. Primera, en el análisis, que ha de realizarse desde la base hacia la punta, la información clínica aumenta también en sentido ascendente, y segunda, para su interpretación se requieren más conocimientos clínicos que estadísticos. La Figura 5 muestra de forma genérica los 5 niveles de información, y la Figura 6 la interpretación concreta de cada uno de los 5 niveles en nuestro ECA ficticio de la Tabla 4.

Tabla 4. Resultados de un ensayo clínico ficticio en HBP

Ensayo clínico en HBP	
Variable de estudio: Incremento del flujo máximo	
Fármaco X	Placebo
p = 0,035	
Q _{max} = 2,8 ml/s	Q _{max} = 0,6 ml/s
Diferencia media = 2,2 ml/s (CI95% = 1,7 - 2,7)	
NT = 12	



FIGURA 5. Pirámide de los niveles de información de un ensayo clínico. Propuesta del autor.

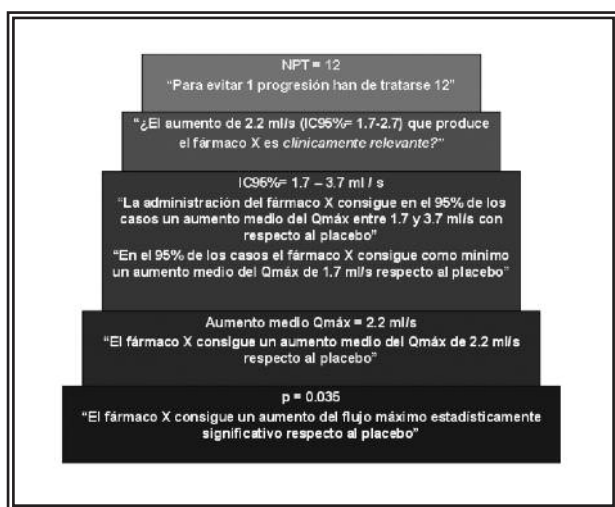


FIGURA 6. Interpretación del ejemplo de la Tabla 4 a partir de la pirámide de los niveles de información.

En primer lugar, ha de analizarse la *significación estadística*, que nos viene dada por el valor p . Ya hemos visto, que cuando es inferior a 0,05 significa que es estadísticamente significativa. En nuestro ejemplo, con un valor de 0,035, lo único que se puede concluir es “El fármaco X consigue un aumento del flujo máximo estadísticamente significativo respecto al placebo”. Sin embargo, de esto no se puede concluir que el fármaco X sea más eficaz o no; lo único que quiere decir es que la probabilidad de que las diferencias observadas entre el fármaco X y el placebo sean debidas al azar es muy baja (3,5%), por lo que podemos aceptar que estas diferencias se deben realmente a la acción del fármaco X. Como ya se ha ido comentando, a lo largo de este

texto, el principal problema a la hora de interpretar un ECA, es que habitualmente, de una forma miope, sólo nos focalizamos en el valor p , pero no analizamos los 4 peldaños superiores, que como veremos a continuación, nos irán proporcionando más información clínica.

A continuación, el segundo paso ha de ser el *análisis del resultado*. En este caso, dado que el fármaco X y el placebo consiguen, respectivamente, un aumento del flujo máximo de 2,8 y 0,6 ml/s, la diferencia media es de 2,2 ml/s. Por ello, lo que se puede decir, tal como recoge la Figura 6, es que “El fármaco X consigue un aumento medio del flujo máximo de 2,2 ml/s respecto al placebo”. Fijese, que ya en este segundo escalón aparecen unidades (ml/s), lo cual es más inteligible para el clínico, que la información aportada en el primer escalón (significación estadística) ya que sólo es un número sin unidades (0,035). Por otro lado, si se asciende al tercer escalón (*Intervalo de Confianza*), se incrementará la información del segundo escalón (Resultado), ya que se nos informa además de que en el 95% de los casos tratados con el fármaco X, el flujo medio aumentará de 1,7 a 3,7 ml/s, o dicho de una forma más conservadora, que en el 95% de los casos el flujo máximo se incrementará como mínimo en 1,7 ml/s. Por cierto, ¿cómo le aportaremos más información clínica al paciente? ¿Diciéndole, que le vamos a tratar con un fármaco que tiene una $p=0,035$ (Primer escalón) o explicándole, que si se toma dicho fármaco, con toda probabilidad su flujo máximo aumentará como mínimo en 1,7 ml/s (Tercer escalón)?

Seguidamente, nos queda valorar los dos últimos escalones de la utilidad clínica en el paciente y en la población. Observe, que para analizar estos dos niveles de información, que son los más importantes, no se necesita ningún conocimiento estadístico, sino tener conocimientos sobre la materia y sentido común. Así, en el escalón de la “*Utilidad Clínica*” deberíamos respondernos a la siguiente pregunta. “¿El aumento medio de 2,2 ml/s de flujo máximo es clínicamente relevante? Quien responderá mejor a esta cuestión no es un experto en estadística, sino el urólogo, porque domina la materia, o incluso, el paciente porque a él le repercutirá o no los supuestos beneficios del tratamiento con el fármaco X. Piense, que en la literatura, en los congresos, o en la información de la Industria Farmacéutica nos podemos encontrar con muchos ejemplos de estudios con una p muy sig-

nificativa, pero cuyo resultado implique una nula o escasa relevancia clínica. Finalmente, ha de analizarse la *utilidad en la población* (último escalón), valorando el coste económico y el NPT. Evidentemente, que a un menor NPT, el fármaco es más eficaz; pero tan importante como valorar el número en sí, es el análisis de lo que se está previniendo. Así en el ejemplo de la HBP, no tiene la misma repercusión considerar como progresión el empeoramiento de la sintomatología que la necesidad de cirugía. Otro ejemplo: puede ser más útil para la población tratar a 25 pacientes de hipercolesterolemia con estatinas para evitar un infarto de miocardio, que a 8 pacientes afectados de HBP con un inhibidor de la 5 alfa-reductasa para evitar una progresión entendida como empeoramiento de la sintomatología.

Finalmente, como mensajes finales, primero, cuando analice un ECA, no se limite a interpretar únicamente la significación estadística, sino que valore también el resultado de la acción del fármaco y la utilidad clínica que tiene en el paciente y en la población general. Y segundo, desconfíe de aquellos trabajos que sólo presentan o se focalizan en la significación estadística.

Correspondencia autor: Dr. Félix Millán
Servicio de Urología. Fundació Puigvert
Cartagena 340-350 - 08025 - Barcelona
Tel: +34 93 4169700 Fax: +34 93 4169730
E-mail autor: fmillan@fundacio-puigvert.es/fmillan@ono.com
Información artículo: Original - Metodología
Trabajo recibido: abril 2009
Trabajo aceptado: mayo 2009