

INTERPRETACIÓN DE RESULTADOS DE ENSAYOS CLÍNICOS EN HIPERPLASIA BENIGNA PROSTÁTICA

F. MILLÁN RODRÍGUEZ*, J.M. RODRÍGUEZ VALLEJO**,
J. CARBALLIDO RODRÍGUEZ***

Servicio de Urología. Fundación Puigvert. Barcelona. **Gerente Senior. Departamento de Marketing. Grupo MSD. Madrid. *Servicio de Urología. Hospital Universitario Clínica Puerta de Hierro. Madrid.*

PALABRAS CLAVE:

Hiperplasia benigna de próstata. Ensayo clínico. Medicina basada en la evidencia.

KEY WORDS:

Benign prostatic hyperplasia. Clinical trial. Evidence based medicine.

Actas Urol Esp. 26 (1): 4-14, 2002

RESUMEN

El ensayo clínico aleatorio (ECA) es el estudio más adecuado para evaluar la eficacia de un tratamiento en hiperplasia benigna de próstata (HBP). Aunque la mayoría de urólogos no colaborarán en el desarrollo de un ECA, sí que van a tratar pacientes afectados de HBP, por lo que tiene gran interés saber discernir si un trabajo sobre un ECA en HBP está bien diseñado y sus conclusiones son correctas. El objetivo de este artículo es proporcionar los elementos básicos de análisis que permitan a un urólogo evaluar la calidad y el grado de evidencia de un ECA en HBP. El trabajo gira entorno de los tres pilares básicos de cualquier ECA: comprobar si el estudio se ha realizado de forma correcta (validez interna), evaluar si el tratamiento consigue mejoría clínica importante (relevancia de los resultados) y la aplicabilidad de los resultados del estudio a nuestros pacientes (validez externa). Como se comprobará para analizar estos elementos no se requieren grandes conocimientos estadísticos sino el uso del juicio clínico y del sentido común.

ABSTRACT

The random clinical trial (RCT) is the most suitable study to evaluate the treatment effectiveness in the benign prostatic hyperplasia (BPH). Although most of the urologists will not collaborate in a RCT development, they will treat BPH patients, so it is very important to know if a CRT in BPH is well designed and their conclusions are correct. The aim of this article is to give the basic elements of analysis that urologists need in order to evaluate the quality and the level of evidence of a RCT in BPH. This article emphasizes the three main elements of a RCT: to check if the study has been correctly performed (internal validity), to evaluate if the treatment achieves an important clinical improvement (relevance of the results) and the applicability of the results in our patients (external validity). The article shows that to analyse these elements common sense and clinical judgment are needed rather than statistical knowledge.

Se disponen de diferentes tipos de estudios para evaluar la acción de una intervención o un tratamiento sobre una enfermedad (Tabla I). 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 tanto en hiperplasia prostática benigna (HBP) como en otras patologías. 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 HBP 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 farmacéutica que presente un ECA sobre HBP?

El objetivo de la presente comunicación es contribuir a revisar los conocimientos necesarios para evaluar e interpretar un ECA en HBP y por extensión en otras patologías. Habitualmente, lo primero que se piensa cuando se ha de abordar el estudio de un problema con 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.

TABLA I
CLASIFICACIÓN DE LOS ESTUDIOS DE MAYOR A MENOR EVIDENCIA CIENTÍFICA

1. Ensayo clínico aleatorio.
2. Ensayo clínico no aleatorio.
3. Estudio de cohortes.
4. Estudio de casos y controles.
5. Estudios transversales.
6. Serie de casos.
7. Caso clínico aislado.

Normalmente, los tratados de estadística o epidemiología enseñan cómo realizar un ECA pero con menos frecuencia cómo interpretarlo. Del mismo modo, cuando estudiamos Literatura ¿qué se nos enseñó? ¿a escribir como Cervantes? ¿o a interpretar y disfrutar leyendo “El Quijote”? Por todo ello, nuestra intención es desarrollar a través de un enfoque práctico los conocimientos mínimos para evaluar un ECA.

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. Pongamos un ejemplo. Si se quiere valorar el efecto del tabaquismo en la aparición del tumor vesical, se debería hacer mediante un estudio observacional. Así se estudiará una cohorte de fumadores y otra de no fumadores y al cabo de un tiempo se contabilizará la diferente incidencia de tumor vesical en ambos grupos. Con el mismo ejemplo, si se quisiera realizar un estudio experimental 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. 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 II). 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.

TABLA II

ELEMENTOS A ANALIZAR EN UN ENSAYO CLÍNICO ALEATORIO

Formulación de la hipótesis <ul style="list-style-type: none"> – ¿El estudio aporta algo nuevo? – ¿El objetivo del estudio está bien definido?
Validez interna <ul style="list-style-type: none"> – Asignación aleatoria de los sujetos. – Seguimiento de los pacientes. – Estrategia de análisis por intención de tratar. – Enmascaramiento. – Comparabilidad inicial de los grupos.
Relevancia de los resultados <ul style="list-style-type: none"> – Precisión. – Magnitud del efecto.
Validez externa <ul style="list-style-type: none"> – Diferencias entre los pacientes seleccionados y la población a tratar. – Modificación del pronóstico. – Relación entre beneficios, seguridad y coste.

1. 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.

Por otro lado, 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 finasterida. En cambio, sí que puede ser novedoso estudiar la eficacia de estos fármacos en indicaciones nuevas. En cual-

quier caso, la existencia de diferentes ECA que tengan el mismo objetivo puede servir posteriormente para la realización de un meta-análisis que valore conjuntamente todos los estudios publicados previamente.

También, hay que recordar que un ECA tiene que tener claramente definido el objetivo de 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. Pongamos, otro 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 ejemplo siguiente:

[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 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 con el 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.

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

En un ECA siempre existen dos hipótesis. La hipótesis nula (H_0) es la que se pretende rechazar y la hipótesis alternativa (H_1) la que se pretende demostrar. Pongamos un ejemplo:

[d] “ H_0 : Los pacientes afectados de HBP tratados mediante el fármaco B no presentan un aumento del flujo máximo urinario estadísticamente significativo respecto a los tratados con placebo”.

[e] “ H_1 : Los pacientes afectados de HBP tratados mediante el fármaco B presentan un aumento del flujo máximo urinario estadísticamente significativo 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 H_1 . 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 H_0 y que por error se haya aceptado la H_1 . Este es el llamado error α (probabilidad de aceptar la H_1 cuando la verdadera es la H_0). Habitualmente se acepta un error de un 5%, por eso siempre se dice que aceptamos la H_1 cuando la $p < 0,05$. 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 sintomática 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 pasa con otras posibles variables desconocidas que puedan influir en el resultado y que no podamos evitar?. Imaginemos que hipotéticamente la expresión de un determinado gen estuviese ligada a un peor resultado en el tratamiento de la HBP, y que los dos grupos estuviesen desequilibrados respecto a esta variable. Hasta que no se descubriera la importancia de este gen, por ejemplo dentro de 10 años, no podríamos sospechar la escasa validez del estudio.

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 lo 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 pueden 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 sintomatológica 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 las características de los pacientes que abandonan como el número que lo hacen.

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 III). Un ECA pretende comparar con 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:

- 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.

TABLA III

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

– 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 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, el único enmascaramiento que se podría llevar a cabo es el “simple ciego” ya que el investigador ha de conocer a la fuerza a qué grupo ha sido asignado cada paciente.

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.

3. 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 poco al clínico:

[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 precisión y la magnitud del efecto.

1. 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”.

No obstante, 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%. 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 expresa 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”.

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 definición [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.

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 ml/s ($IC_{95\%} = 1,9-2,3$) se habría obtenido por ejemplo 2,1 ml/s (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.

2. 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, como hace el intervalo de confianza. 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 IV). 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 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 (2,7-3,9), mientras que con el

TABLA IV
SIGNIFICACIÓN ESTADÍSTICA Y MAGNITUD DEL EFECTO

Alfabloqueante	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	$P = 0,00001$	2,5	1,8-3,2

alfabloqueante Y el aumento medio del flujo máximo es menor 2,5 ml/s (1,8-3,2). 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.

Llegados a este punto, hemos ido subiendo tres peldaños: el valor “p”, el valor del resultado y el valor del intervalo de confianza (Tabla V). 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 ser de 2 ó 3 ml/s. Ante este hecho, los urólogos debemos plantearnos ¿el hecho de que un paciente aumente su flujo máximo en 2 o incluso 3 ml/s lo va a notar en su forma de orinar? Es más, ¿esta diferencia de 2 ó 3 ml/s es clínicamente relevante cuando está demostrado

TABLA V
NIVELES DE INFORMACIÓN DE UN ESTUDIO

	Significación estadística
I	$P = 0,035$ “El fármaco X consigue un aumento del flujo máximo estadísticamente significativo respecto al placebo”
II	Resultado
	Aumento medio del $Q_{\max} = 3,2$ ml/s “El fármaco X consigue un aumento medio del flujo máximo de 3,2 ml/s respecto al placebo”
III	Intervalo de confianza 95% del resultado
	$IC_{95\%} = 2,7 - 3,7$ ml/s “La administración del fármaco X consigue en el 95% de los casos un aumento medio del flujo máximo entre 2,7 y 3,7 ml/s con respecto al placebo” “En el 95% de los casos el fármaco X consigue como mínimo un aumento medio del flujo máximo de 2,7 ml/s respecto al placebo”
IV	Utilidad clínica
	“¿El aumento de 2,7 a 3,7 ml/s en el flujo máximo que produce el fármaco X es clínicamente relevante?”

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 son muchas las variables que pueden 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.

Resumiendo, 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. A continuación, ha de valorarse la repercusión de estos resultados aplicados a la práctica clínica diaria. Se ha de ser cauteloso con aquellos trabajos que únicamente remarcan que es estadísticamente significativo pero que no informan o lo hacen de pasada sobre el resultado de la prueba, su precisión o su relevancia clínica.

4. 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 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 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 esta características el resultado podría haber sido estadísticamente diferente.

¿Cuál es el problema que nos encontramos habitualmente con los ECA de HBP 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 (Tabla VI).

TABLA VI

DIFERENCIAS ENTRE LA MUESTRA ESTUDIADA Y LA POBLACIÓN A TRATAR

Características típicas de los pacientes de un ensayo clínico en HBP	Características típicas de los pacientes en la práctica clínica diaria
<ul style="list-style-type: none"> – Edad entre 50 y 70 años. – Próstata de pequeño tamaño. – Ausencia de infección urinaria. – Ausencia de diabetes. – Ausencia de enfermedades neurológicas. – Valores bajos de PSA. – Obstrucción infravesical leve. 	<ul style="list-style-type: none"> – Edad avanzada. – Próstatas de todos los tamaños. – Con frecuencia riesgo quirúrgico elevado. – Frecuentemente asociación a infección urinaria, diabetes, enfermedades neurológicas o toma de Sintrom®. – En ocasiones, asociación a cáncer de próstata. – Frecuentemente obstrucción infravesical moderada o severa.

Habitualmente se incluyen pacientes entre 50 y 70 años, con próstatas de pequeño tamaño, sin infección asociada, sin diabetes, sin enfermedades neurológicas, con valores de PSA bajos, con obstrucción leve... Mientras que en la realidad diaria encontramos pacientes ASA 4 (y porque no hay 5), 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 (Tabla VII), 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 fármaco 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*.

TABLA VII

MODIFICACIÓN DEL PRONÓSTICO DE LA ENFERMEDAD

Grupo de tratamiento	Aumento del flujo máximo a los 3 meses de tratamiento (ml/s)	Intervención por retención urinaria al año (%)
Fármaco Z	3,4 (2,5-4,3)	15
Placebo	1,5 (0,8-2,2)	21

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 ECA 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, 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.

EPÍLOGO

La lectura de este trabajo podría dejar la sensación de que el diseño y la ejecución de un ECA es una carrera de obstáculos. Lo que realmente refleja es el precio que se ha de pagar para alcanzar un alto nivel de evidencia científica que garantice la fiabilidad de los resultados obtenidos, para poder ser aplicados en nuestra práctica clínica diaria. En cualquier caso, debe quedar bien establecido que el objetivo no era cómo confeccionar un ECA sino cómo leerlo e interpretarlo. Esperamos que después de su lectura, cuando el lector tenga ante sí un ECA, le pierda el miedo a su interpretación y haya adquirido los conocimientos necesarios para poder evaluarlo.

REFERENCIAS

1. SACKETT DL, RICHARDSON WS, ROSENBERG W, HAYNES RB: Medicina basada en la evidencia. Cómo ejercer y enseñar la MBE. Ed. Churchill-Livingstone. Madrid, 1997.
2. HULLEY SB, CUMMINGS SR: Diseño de la investigación clínica. Un enfoque epidemiológico. Ed. Doyma. Barcelona, 1993.
3. FLETCHER RH, FLETCHER SW, WAGNER EH: Epidemiología clínica. Aspectos fundamentales. Ed. Masson-Williams & Wilkins. Barcelona, 1998.
4. MATTHEWS DE, FAREWELL VT: Using and understanding medical statistics. Ed. Karger. Basel, 1996.
5. ROTHMAN KJ, GREENLAND S: Modern epidemiology. Ed. Lippincott Williams & Wilkins. Filadelfia, 1998.
6. CARBALLIDO RODRÍGUEZ J: El ensayo clínico en Urología. Ed. Ergon. Madrid, 1997.
7. CARBALLIDO RODRÍGUEZ J: El ensayo clínico en hiperplasia prostática benigna. ¿Dónde estamos?. *Rev Clin Esp* 1999; **199**: 48-57.
8. CARBALLIDO RODRÍGUEZ J, RODRÍGUEZ VALLEJO JM, DEL LLANO SEÑARIS J: Hiperplasia prostática benigna y medicina basada en la evidencia: su aproximación a la práctica benigna. *Med Clin* 2000; **114**: 96-104.
9. World Health Organization. 4th International consulting on benign prostatic hyperplasia. *Proceedings*. Paris, 1997.

Dr. F. Millán Rodríguez
Servicio de Urología. Fundación Puigvert
C/ Cartagena, 340-350 - 08025 Barcelona

(Trabajo recibido el 13 julio de 2001)