

FUNDADOS POR E. PEREZ CASTRO, A. PUIGVERT GORRO Y  
L. CIFUENTES DELATTE

Director / Editor: E. Pérez-Castro Ellendt  
Editor Asociado: L. Martínez-Piñeiro Lorenzo

TOMO 59  
Número 5  
Junio 2006

## LA TIRANIA DE LA p

Antonio Berenguer. Álvaro Paez Bordes<sup>1</sup> y Marcos Luján Jalem.

Servicio de Urología. Hospital de Getafe y Servicio de Urología. Hospital de Fuenlabrada<sup>1</sup>. Madrid. España.

**E**n la actualidad, no es imaginable un trabajo científico que no vaya acompañado de un *adecuado* tratamiento estadístico. En este sentido, existe una total subordinación a los criterios de *adecuación* estadística definidos a principios del siglo XX: el famoso –y arbitrario– 5% establecido como límite para la probabilidad de que la diferencia entre dos grupos obedezca al azar, todavía constituye el principal indicador de relevancia científica. Con gran frecuencia, esa aproximación a la *adecuación* metodológica es la única condición para que un trabajo científico sea aceptado en una revista especializada o un proyecto de investigación sea financiado. En Estados Unidos, un producto farmacéutico puede ser anunciado como efectivo siempre y cuando la diferencia de efecto en relación al placebo resulte estadísticamente significativa ( $p < 0,05$ ), (1) con independencia de que su efecto sea 2 o 200 veces superior al del placebo.

Como es sabido, la determinación del **tamaño muestral** constituye un pieza determinante en el diseño de un estudio. Así, diferencias irrelevantes entre dos formas de tratamiento (o entre dos procedimientos diagnósticos) desde el punto de vista estadístico, adquieren relevancia (su  $p$  alcanza un valor  $< 0,05$ ) en el momento en que el mismo estudio es dotado de suficiente tamaño muestral. Para comprobar este último extremo, invitamos al lector a hacer uso de uno de los innumerables programas para el cálculo del tamaño muestral; en primer lugar, trasládese a la siguiente página web de acceso libre: <http://calculators.stat.ucla.edu/powercalc/normal/n-2-unequal/>; (2) en la casilla correspondiente a  $\mu_1$  (*The Mean of Population 1*) teclee 2046 (el peso de los recién nacidos en el brazo de tratamiento activo en un reciente estudio que explora el efecto preventivo de un tratamiento

## EDITORIAL

antibiótico sobre el nacimiento pretérmino). (3) A continuación, introduzca 2464 en la casilla correspondiente a  $\mu_2$  (el peso de los recién nacidos en el brazo de tratamiento placebo en el mismo estudio); acto seguido, teclee 65 en la casilla  $N_1$  (número de embarazadas incluidas en el brazo placebo) y 59 en la casilla  $N_2$  (número de embarazadas incluidas en el brazo de tratamiento). A continuación teclee 1209 en la casilla **Sigma 1** (desviación estándar del peso en el grupo de tratamiento activo) y 1067 en la casilla **Sigma 2** (desviación estándar del peso en el grupo placebo). En la casilla **Significance Level** teclee ?. Finalmente, introduzca .80 (un punto seguido de 80) en la casilla **Power** (poder estadístico más habitualmente utilizado en biomedicina) y seleccione **2 sides** (2 colas). Pulse **Calculate** y obtendrá el nivel de significación para esa comparación: el resultado es 0.231, esto es,  $p=0.231$ , que, por cierto, no coincide con el publicado por los autores del documento. Pruebe ahora a sustituir los números de embarazadas en cada brazo (teclea 160 en las casillas  $N_1$  y  $N_2$ , por ejemplo); mantenga el resto de parámetros inmodificados y pulse **Calculate**: en efecto, el nivel de significación ha variado, de manera que la diferencia entre ambas formas de tratamiento adquiere ya la mágica significación estadística. En definitiva, la probabilidad de detectar una diferencia entre dos formas de tratamiento depende excesivamente del tamaño muestral, lo que convierte a la significación estadística en un indicador de relevancia clínica muy vulnerable; es el momento de **medir el efecto** del tratamiento.

Pruebe ahora a calcular el número de individuos necesario tratar (NNT) con finasteride para evitar una progresión (entendida como el incremento del IPSS en 4 o más puntos, o el desarrollo de incontinencia urinaria, insuficiencia renal o infecciones urinarias de repetición) en pacientes con próstatas de tamaño igual o superior a 40 ml. (4) El riesgo de progresión en el grupo placebo de ese estudio es del 7,34% (intervalo de confianza para el 95% de las observaciones, 5,55–9,53), mientras que ese riesgo alcanza el 1,77% (intervalo de confianza para el 95% de las observaciones, 1,05–2,79) en el grupo de pacientes tratados con finasteride. La diferencia en el riesgo de progresión en términos absolutos (reducción del riesgo absoluto, RRA) asciende al 5,57% ( $7,34\% - 1,77\% = 5,57\%$ ). Dividiendo 100 entre la RRA se obtiene el NNT; en este caso, en efecto, el NNT es de aproximadamente 18 ( $100/5,57=17,95$ ). Ahora, ¿considera que *merece la pena* el tratamiento a la vista del NNT? Decidir si algo *merece o no la pena* cuando la salud o la vida están de por medio puede parecer frívolo; sin embargo, la consideración del NNT y de la frecuencia de efectos adversos (en términos de número de individuos necesario tratar para sufrir un efecto adverso, o NNH, del inglés *number necessary to harm*) puede permitir una decisión ajustada. Se trata, en definitiva, de introducir las medidas de efecto (el NNT es una de las **medidas de efecto** más populares) como guía para la práctica clínica. Su principal virtud reside en que la cuantificación del efecto es *invulnerable* a las modificaciones del tamaño muestral, lo que lo convierte en un indicador más robusto para el clínico.

Finalmente, la consideración de la significación estadística y de las medidas de efecto resulta inapropiada si no se tienen en cuenta los **cambios clínicos** que introduce un determinado tratamiento. En otras palabras, diferencias estadísticamente significativas en la efectividad de un tratamiento pueden asociarse (o no hacerlo) con diferencias sustanciales

## EDITORIAL

(o irrelevantes) en el efecto de esos tratamientos. Sin embargo, es la valoración que hace el paciente (su consideración de la mejoría que se asocia con una forma de tratamiento, con independencia del grado de significación estadística alcanzado en los estudios comparativos, y del NNT de ese tratamiento) la que establece, finalmente, la verdadera adecuación de un tratamiento. Existen múltiples alternativas para medir el **cambio clínico**.

En suma, la sola consideración del nivel de significación estadística a la hora de optar por una forma de tratamiento puede resultar poco adecuada. La interpretación de los hallazgos de la investigación biomédica debería de tener lugar en el contexto del análisis de la magnitud de los cambios de efecto y de la consideración de los cambios clínicos.

### BIBLIOGRAFÍA y

#### LECTURAS RECOMENDADAS

1. ZEEV, N. Kain. The Legend of the P Value. *Anesth Analg*;101:1454, 2005.
2. <http://calculators.stat.ucla.edu/powercalc/normal/n-2-unequal/> Ultimo acceso, 16/03/2006
3. ANDREWS, W.; GOLDENBERG, R.; HAUTH, J. y cols.: "Interconceptional antibiotics to prevent spontaneous preterm birth: A randomized clinical trial". *Am J Obstet Gynecol*; 194:617, 2006.
4. KAPLAN, S.; McCONNELL, J.D.; ROEHRBORN, C.G. y cols.: "Combination Therapy With Doxazosin and Finasteride for Benign Prostatic Hyperplasia in Patients With Lower Urinary Tract Symptoms and a Baseline Total Prostate Volume of 25 ml or Greater". *J Urol*; 175:217, 2006.