Artículo especial

Localizador web
Artículo 277.460

Recomendaciones para la valoración clínica de los resultados en literatura biomédica

Javier Escrig-Sos, David Martínez-Ramos, Carmen Villegas-Cánovas, Juan Manuel Miralles-Tena, Isabel Rivadulla-Serrano y José María Daroca-José
Servicio de Cirugía. Hospital General de Castellón. Castellón de la Plana. Castellón. España.

La valoración y la interpretación de los resultados de un estudio clínico son un auténtico reto para el profesional médico. En el presente artículo se ofrecen las bases generales para una valoración crítica y comedida, partiendo de aspectos fundamentales del diseño y de la estadística, así como de la aplicación de los resultados a nuestros propios pacientes según criterios de riesgo y beneficio. Se hace hincapié en los principales errores y en las trampas que se debe evitar.

Palabras clave: *Diseño explicativo. Diseño pragmático. Balance riesgo-beneficio.*

RECOMMENDATIONS FOR THE CLINICAL EVALUATION OF RESULTS IN THE BIOMEDICAL LITERATURE

The assessment and interpretation of the results of a clinical study are a real challenge for the clinicians. In this paper we establish a general basis for a critical and reserved assessment of these, from the fundamental aspects of the design and statistics, as well as the application of the results to our own patients according to risk and benefit criteria. Main errors and the traps that should be avoided are emphasised.

Key words: *Explanatory design. Pragmatic design. Risk-Benefit analysis.*

Introducción

Mucho se ha escrito en revistas biomédicas, así como en gruesos tomos de papel impreso, sobre la valoración metodológica y la interpretación de los resultados estadísticos de los artículos de investigación clínica. Pero a juzgar por lo percibido en el día a día entre los colegas y por lo que se puede leer en las propias revistas médicas, uno tiene el pálpito de que todavía se hace necesario insistir en el tema. Generalmente, la divulgación metodológica y estadística se hace en sentido positivo, es decir, en lo que hay que hacer, pero es tan grande este campo que no siempre es fácil comunicar rotundamente ciertos conceptos. Por ello, aquí y con el fin de destacar algunas ideas básicas no siempre tenidas en cuenta, tomaremos la vía opuesta: lo que no hay que hacer pero se hace o, lo que es lo mismo, las más esenciales precauciones que hay que tomar cuando se leen números y más números, dependiendo del tipo de artículo que esté en nuestras manos.

Tipo de estudio

La estadística sólo es una herramienta de análisis, y por ello está en función de lo que se quiere analizar y del camino que se decide tomar para tal cometido. Su uso y su interpretación, por lo tanto, dependen del diseño del estudio. Aunque hay muchas clasificaciones referentes al diseño de investigación, quizá la que más interesa al clínico para la posterior valoración del alcance de los resultados sea, curiosamente, una de índole muy general utilizada para catalogar los ensayos clínicos: la que divide a los estudios en explicativos y pragmáticos[1]. Este propósito tan fundamental no siempre se aclara de forma meridiana en los objetivos de los artículos; consiguientemente, es lo primero que los propios lectores deben descubrir porque, como veremos a continuación, es una cuestión de mucho peso.

Un estudio pragmático está dirigido, ni más ni menos, hacia una toma de decisión, y esto lleva necesariamente a concluir que un resultado es mejor (o peor) que otro y, desde ahí, que es conveniente abandonar una práctica determinada para adoptar la otra que aparenta ser mejor. Como cualquier decisión de este tipo puede ser muy deli-

Correspondencia: Dr. J. Escrig Sos.
Servicio de Cirugía. Hospital General de Castellón.
Avda. Benicàssim, s/n. 12004 Castellón de la Plana. Castellón. España.
Correo electrónico: escrig_vicsos@gva.es

Manuscrito recibido el 24-10-2007 y aceptado el 31-1-2008.

cada en el plano real, es ineludible exigir al trabajo ciertas condiciones:

1. Es necesario que conste el cálculo del tamaño de la muestra. Ante todo, porque si hay oportunidad de demostrar, con buen fundamento probabilístico, una diferencia interesante, esta oportunidad debe aprovecharse, y para ello es necesario contar con suficientes sujetos de análisis. Después ocurrirá lo que ocurra, pero no hay que arriesgarse demasiado a perder tiempo, dinero y esfuerzos.

2. Para ello hay que decidir qué es una diferencia suficiente para tomar una decisión que conlleve un cambio en la práctica diaria. Algunos lo llaman MCID *(minimal clinically important difference)* y en los últimos años este concepto está muy presente en artículos que versan sobre metodología de investigación clínica[2-4]. Definir un valor concreto para esa diferencia es ciertamente difícil y siempre puede estar sujeto a polémica, pero es necesario que los autores lo fijen claramente y, por nuestra parte, debemos detenernos a pensar unos minutos si lo aceptamos como bueno, de acuerdo con nuestros conocimientos. Nunca aceptemos este dato pasivamente. Algunos cometen la trampa de hacer el análisis estadístico con la muestra de que disponen en un momento dado, porque han comprobado que en ese momento existen diferencias estadísticamente significativas; entonces adaptan esa diferencia o una similar como supuesto punto de partida para que los números cuadren. En ese entramado, la diferencia suele ser superior, o muy superior, a la mínima diferencia que ya podría tener trascendencia clínica. Rara vez es al contrario. Precisamente la mínima diferencia que fuera claramente trascendental es de la que se debería partir para calcular la muestra[5]. Sospechen, pues, de cálculos de muestra basados en diferencias muy amplias desde el punto de vista clínico.

3. Aceptado como correcto el número necesario de sujetos que era necesario reclutar, hemos de comprobar si el tipo de sujetos escrutados es de lo más variopinto posible dentro de la enfermedad de que se trate. Una toma de decisión afectará a todos los implicados en el problema, a amplias capas de la población, y no puede haber ninguna gran selección en este colectivo. En otras palabras, un estudio pragmático debe reclutar a todos los pacientes que acudan con el problema. Así, en el contexto general de un esquema de tratamiento citostático de un cáncer con metástasis, un paciente podría perfectamente ser incluido en la investigación ante la primera comprobación por imagen, en principio, sin necesidad estricta de un escrutinio de toda la economía en busca de más metástasis o de una confirmación por biopsia ni de ninguna selección según el tratamiento inicial recibido, etc. Los criterios de exclusión no pueden ser, pues, muy amplios.

4. En un estudio pragmático la comparación de un nuevo procedimiento debe realizarse con otro que hasta ese momento constituya la "mejor práctica estándar". Parece una obviedad, pero en realidad no lo es, puesto que no siempre se respeta este principio. Esto, además, exige de los autores una vigilancia muy activa de este grupo control. Los lectores deben cerciorarse de que los resultados en ese grupo sean precisamente los estándar, y no peores de lo que se podría esperar o sea bien conocido.

5. Un estudio pragmático no puede apoyarse en unas mediciones a base de variables llamadas duras, sino que debe basarse en variables de índole muy general, como serían las dependientes del tiempo (supervivencia, intervalo libre de enfermedad, etc.) o de calidad de vida. Así, podría tratarse de un truco si, para justificar un cambio de decisión acerca del resultado de un nuevo tratamiento oncológico en un cáncer metastásico, un estudio se basara solamente en la reducción en milímetros del tamaño de una metástasis. Estas variables tan duras tienen tendencia a mostrar diferencias significativas con mayor facilidad, y ahí radica la posible trampa, además de que por sí solas no informan de las repercusiones finales a que puedan conducir.

6. Al analizar los resultados, en un estudio pragmático conviene aplicar el principio de intención de tratar. Si un sujeto, por razones de necesidad, fue finalmente tratado en el grupo opuesto al que se le asignó aleatoriamente, el resultado debe endosarse siempre al grupo al que fue inicialmente asignado. Esto es así porque lo que se pretende en un estudio de este tipo es resaltar la relevancia clínica de la intención del tratamiento, incluidos sus avatares, más que el resultado concreto en situación ideal.

7. El archifamoso corte de significación estadística ($p < 0,05$), tan recurrido para resaltar la importancia de un resultado en cualquier tipo de estudio, en realidad y de forma categórica sólo tiene congruencia si hay que tomar una decisión. Recurrir fuera de este contexto a la ley del todo o nada para juzgar un resultado supone forzar la interpretación del valor p de una forma carente de todo sentido científico[6]. Los intervalos de confianza referidos a los resultados principales dan una información muy valiosa sobre la importancia real del resultado y son un complemento imprescindible del valor p[7]. Como puede desprenderse de lo afirmado hasta ahora, un estudio enfocado a una toma de decisión sólo puede corresponder a un ensayo clínico en el que se comparan dos actuaciones, una nueva frente a otra ya establecida. Desechen, pues, cualquier alusión a la significación estadística interpretada según la ley del todo o nada, sea cual sea su intención, fuera de este ámbito.

Frente al sentido pragmático, un estudio explicativo busca simplemente investigar, ampliar conocimientos e informar de las relaciones existentes entre ciertas variables, sin que ello presuponga la necesidad automática o inmediata de una toma de decisión. Aquí no sólo caben ciertos ensayos clínicos: por extensión, se puede reconocer como tales desde las series de casos hasta los estudios de casos y controles y los de cohortes. Cuando se trata de un ensayo clínico, el grupo control está constituido por un placebo, porque en un ensayo clínico explicativo lo que se pretende es valorar la eficacia pura y dura de una intervención médica o quirúrgica, o sus mecanismos de acción y los potenciales beneficios. Estos beneficios potenciales siempre se deben comprobar después en un estudio pragmático. Realmente, las condiciones exigibles a esta clase de investigación no son más laxas que antes, sino que cambian en algunos matices:

1. Está claro que, si se trata de un ensayo clínico, hay que partir del cálculo del tamaño de la muestra según los

mismos presupuestos que antes, anteponiendo a todo la importancia clínica. En una serie de casos, puede no ser formalmente necesario llevar a cabo este cálculo, pero es preciso que los autores justifiquen por qué eligen un periodo de reclutamiento determinado y no otro. Sin embargo, rara vez hay comentarios en tal sentido en esta clase de artículos, cuando en verdad son absolutamente exigibles.

2. Aquí, no obstante, la muestra puede estar más dirigida o seleccionada. En el caso de la eficacia en la reducción de una metástasis de un tumor, podría ser lícito elegir a sujetos con una tumoración de cierto tamaño cuya reducción fuera fácilmente mensurable. Si se trata de una operación quirúrgica, se podría elegir, en términos explicativos, a sujetos claramente operables y resecables, para facilitar un tratamiento uniforme sin desviaciones del protocolo.

3. Las condiciones de un estudio explicativo remedan las de laboratorio. La definición y la elección de las variables de entorno y su control deben ser exhaustivas, así como la recogida de datos. De alguna manera, esto ha de describirse en el texto del artículo, por lo que es fácil deducir que hay que desconfiar de los que no lo hacen. Dicho de otra forma, ha de existir un protocolo muy estricto y se debe cumplir a rajatabla.

4. Las variables de resultado serán, generalmente, las que antes se han definido como duras: o una medida cuantitativa concreta o una cualidad tajante y excluyente de las demás. Sirva el ejemplo de la medida en milímetros de un tumor o una estancia postoperatoria. La variable más dura que existe en medicina es la que no ofrece dudas de interpretación ni de medida: muerto o vivo. Sin embargo, no lo es tanto un tiempo de supervivencia, puesto que en los que no han muerto al final de un estudio su verdadero valor queda en suspense.

5. Si un estudio explicativo recuerda las condiciones de laboratorio, no caben alteraciones en su protocolo. Aquí no tiene indicación alguna el análisis por intención de tratar. Las pérdidas son lamentables y no deberían ser muchas. En cualquier caso, desconfíen de un estudio, tanto explicativo como pragmático, que no detalle las pérdidas de pacientes habidas en su desarrollo y que no incluya algún dato estadístico sobre sus características.

6. La aplicación del concepto de significación estadística de un resultado a partir de un valor de corte concreto (generalmente, $p < 0,05$) en un estudio explicativo carece de todo sentido. Para explicar un vínculo entre dos variables en la población, tan plausible es que exista relación si el valor p en la muestra es 0,03 como si es 0,07 y quizá también si vale 0,11. Hablar de significación estadística en este contexto explicativo como si de tomar una decisión se tratara frena la posibilidad de la discusión y el razonamiento científico sobre la realidad de las cosas y las posibles consecuencias de esa relación y, a la postre, lo que favorece es la pereza intelectual, además de entorpecer la posibilidad de formular interesantes hipótesis susceptibles de comprobación en futuros estudios. De nuevo, los intervalos de confianza son un buen antídoto contra este error[7].

Naturalmente, esta distinción tan absoluta entre lo explicativo y lo pragmático no siempre es posible. ¿Cómo catalogar una comparación entre técnicas quirúrgicas con respecto a la recidiva local de un cáncer de recto? Posiblemente, esto comparta aspectos de ambos mientras no se sepa exactamente la repercusión última de una recidiva local. El consejo general cuando se evalúe un estudio y cuando se interprete la significación estadística del resultado principal en estos diseños "intermedios" es que se tome como un estudio explicativo típico, sobre todo en lo referente a la interpretación de los valores p, hasta que aparezca otro estudio de intención claramente pragmática, que oriente hacia una posible toma de decisión mediante el uso de variables menos duras, como sería la supervivencia general entre los casos operados con una técnica u otra.

Factores de confusión

Si un cirujano operara un tumor sólo en estadios I y II y otro cirujano de igual capacitación sólo operara el mismo tumor en estadios III y IV, es lógico imaginar que la diferencia en supervivencia entre ambos sería muy significativa. En este ejemplo, el estadio tumoral está asociado estadísticamente a la variable referida a los cirujanos, puesto que hay diferencias llamativas entre ambos, y también lo estará a la supervivencia, como siempre sucede con el estadio. Además, el estadio del tumor está cronológicamente (es anterior) fuera de la cadena de hechos que van desde la intervención hasta la supervivencia del paciente. Tiene así todos los requisitos para poder actuar como un factor de confusión. Pues bien, en un análisis estratificado o en un análisis multivariable tipo regresión de Cox en el que se incluyera la variable cirujano y la variable estadio, seguramente la variable cirujano perdería su significación estadística con respecto al resultado de supervivencia, y el estadio mantendría tal significación. El estadio tumoral actúa así como una variable de confusión que convierte en falsa (o espuria) una asociación que en principio parecía clara en el análisis aislado entre cirujano y supervivencia. En un estudio no aleatorizado pueden haber muchas otras variables como el estadio tumoral que tengan tales efectos de confusión. Por ello, aunque se comparen dos tratamientos fuera de un ensayo clínico aleatorizado, un estudio no puede estar dirigido a una toma de decisión efectiva, aunque lo parezca por las trazas de su resultado principal, o sus autores lo hagan entrever en los objetivos, y hay que considerarlo como meramente explicativo, con las correspondientes consecuencias ya comentadas a la hora de interpretar la significación estadística.

Lo que en la práctica hace la aleatorización es distribuir homogéneamente estos factores de confusión entre los grupos que conforman la variable experimental, tanto los que sean conocidos por los investigadores como aquellos de los que no sabemos siquiera que pudieran existir, de modo que no influyan en la comparación principal. Es cierto que las pruebas estadísticas multivariables pueden controlar en gran parte estos efectos de confusión[8], pero tienen ciertas limitaciones matemáticas y, además, sólo pueden ejercer ese control con las variables que conocemos y, por lo tanto, han sido recogidas en la base de datos. La solución no es perfecta, pero es

posible que en muchos estudios no aleatorizados constituya un mínimo absolutamente necesario. Esto es más cierto aún si tenemos en cuenta que descubrir una falsa asociación estadística es trascendental si, en el fondo o a las claras, lo que se pretende es establecer una relación de causalidad entre variables. Como es universalmente sabido, la asociación estadística es imprescindible en una relación de causalidad, aunque después se precisen otros requisitos para establecerla definitivamente.

Muchas variables que reflejan las características basales de la muestra –como edad, comorbilidad, sexo o raza, el propio estadio tumoral, etc., es decir, aquello que el paciente "se trae de casa" y no es modificable por su médico– pueden ser factores de confusión y requieren algún método de control en cuanto al análisis que constituye el objetivo principal de un estudio no aleatorizado. Una auténtica trampa para los lectores, o autoengaño para los autores, muy extendida en estudios tanto prospectivos no aleatorizados como retrospectivos, de la que nos debemos defender a ultranza, consiste en confeccionar una tabla, la "tabla 1" en muchos artículos, en la que estos rasgos basales, hipotéticos factores de confusión, son analizados con respecto a la variable experimental utilizando pruebas de hipótesis. Ante la falta de significación estadística, se da el veredicto de que se distribuyen homogéneamente entre los grupos de la variable experimental, así que no encontrar aquí diferencias estadísticamente significativas es algo así como un buen sustituto de una aleatorización que no se ha llevado a cabo. Nada más falso. Esta falta de significación estadística no cubre que estos factores de confusión produzcan sus temibles consecuencias, ni tampoco previene de los efectos aditivos o multiplicativos entre ellos (efectos de interacción), si es que los hay. Es necesario insistir aquí en que el proceso de aleatorización puede no ser suficiente defensa para la confusión si no hay un tamaño de muestra también suficientemente grande. Desconfíen, pues, de ensayos clínicos escuálidos de muestra y en los que no se plantee alguna forma de ajuste para variables sospechosas de ser factores de confusión.

Comparaciones múltiples

Éste es un asunto áspero y controvertido, del que se han escrito incluso libros enteros, con alto riesgo de que el lector de un artículo las malinterprete y también alto riesgo de error por parte de los autores, alguna que otra vez no exento de cierta intención aviesa. En efecto, no es raro que al comparar, por ejemplo, dos tratamientos quirúrgicos, los resultados se enfoquen desde diversas ópticas para abarcar todos los aspectos posibles de la cuestión. Así, pueden compararse efectos beneficiosos, complicaciones, estancias, pérdidas hemáticas, uso de analgésicos, etc., cada comparación con su valor p. Aquí no suele haber mala intención, como tampoco la hay si se comparara la reducción del tamaño de un tumor tras un tratamiento con citostáticos al mes, a los 3 meses, a los 6 meses, etc., de nuevo cada comparación con su valor p. Donde ya no podemos estar tan seguros de que no haya cierta perversidad es cuando, tras un resultado no significativo en el aspecto fundamental de la investigación, los autores se lanzan "a la caza de la significación estadística", consistente en hacer el mismo análisis pero por subgrupos, como por edad, sexo, estadio tumoral, etc., hasta tropezar con una p significativa. Hoy esto resulta muy fácil de conseguir con los programas estadísticos de ordenador. En un artículo ya antiguo pero absolutamente en vigencia y muy recomendable para su lectura, Mills[9] afirma que, si los datos de un estudio se analizan con muchos enfoques distintos y con la suficiente intensidad hasta torturarlos, acaban dando los resultados que al investigador le hubiese gustado obtener, aunque tales resultados suelen ser científicamente nefastos.

Sin embargo, un valor p sólo es válido para la comparación principal objeto de estudio y para la cual se calculó debidamente el tamaño de la muestra. O para algún análisis de subgrupos siempre que hubiera constado en el diseño antes de realizar cualquier cálculo. Cuando se hacen pruebas de significación en cascada alrededor de un mismo problema de investigación, sin haber sido previstas en el diseño inicial, se corre el riesgo de aumentar el error aleatorio, con lo que p = 0,05 en realidad puede corresponder a un valor más alto. A pesar de lo mucho hablado y escrito sobre el problema, estimamos que al clínico no experto en disquisiciones matemáticas o epistemológicas, sólo le cabe adoptar una de las tres posturas siguientes[10]:

1. En un extremo, admitir sólo como fiable el valor p referido a la pregunta principal de la investigación, y declarar improcedentes por inseguros todos los análisis secundarios no especificados en el diseño inicial del trabajo. Desde luego, los estudios con sólo una cuestión para investigar son los más sólidos y fiables, pero no es menos cierto que así perdemos oportunidades de captar información secundaria que, si bien no es del todo fiel, no por eso puede dejar de ser interesante. Muchos descubrimientos se han hecho por pura casualidad, y con esta postura extrema estaremos impidiendo que esto se pueda producir o vislumbrar.

2. En el extremo opuesto, podemos dar por buenos los valores p de comparaciones múltiples, pero una vez ajustados debidamente en su magnitud. Bonferroni ideó una forma sencilla y directa de ajustar varios valores p: multiplicar cada uno por el número de comparaciones realizadas sobre la misma cuestión. Si hemos hecho 5 comparaciones en serie, una p aislada deberá ser, pues, menor de 0,01 para considerarla realmente significativa para el clásico nivel de 0,05. Hay más modos de hacer ajustes, pero éste es el más sencillo y seguro cuando estamos leyendo un artículo. En realidad el método de Bonferroni es algo conservador, es decir, nos cubre casi demasiado bien del falso positivo (el valor p en realidad es una probabilidad de falso positivo), pero nos descubre algo más de lo debido ante el falso negativo y, a partir de 10 comparaciones relacionadas, algunos dicen que no funciona correctamente[11]. Así pues, ésta es una postura útil, pero tampoco es la ideal.

3. Si en el término medio está la virtud, tal como postula Aristóteles en su *Ética Nicomaquea,* podemos ir valorando los valores p según un orden de importancia: de entrada se da por bueno el referido a la cuestión principal de investigación. Después, podemos hacer el ajuste de

Bonferroni para unas pocas (no muchas) comparaciones secundarias que pudiera haber en el artículo. Si hay comparaciones ya de tercer orden o más en importancia, y alguna de ellas resultara ser estadísticamente significativa, no hagamos ningún ajuste, pero tampoco la tomemos como una cuestión demostrada, sino solamente como una hipótesis para futuras investigaciones o como prueba piloto, siempre que tenga el interés clínico suficiente. Se trata así de alternar una postura de toma de decisión junto con otras de perspectiva más exploradora.

En congresos y reuniones no es infrecuente que los especialistas jóvenes se entrenen en el arte de comunicar utilizando análisis preliminares de investigaciones en curso, aún no totalmente finiquitadas. El problema es distinto, pero los riesgos son similares a los anteriores. Las pruebas estadísticas estándar no arrojan resultados fiables en este contexto, a menos que se utilicen pruebas especiales diseñadas específicamente para tales circunstancias. Por lo tanto, los resultados sólo tendrán valor descriptivo o, a lo sumo, exploratorio, pero de ellos no se puede extrapolar ninguna conclusión. Por lo general, lo que ocurre al finalizar el estudio es que las diferencias no suelen ser tan amplias como en los análisis preliminares, bien porque las cosas suceden muchas veces así en el plano real, bien porque el autor es demasiado artero y ha aprovechado un momento propicio dentro de las oscilaciones que por azar se producen siempre a lo largo de un estudio.

Balance riesgo-beneficio

Aunque un estudio pragmático sea correcto en su diseño, su desarrollo y su posterior análisis, nadie puede negar que sobre el papel impreso se acabe haciendo mucho más hincapié en un hallazgo positivo relacionado con los beneficios que en las complicaciones acaecidas, tanto al leer como al escribir. Seguramente, se comunica sistemáticamente las complicaciones, pero de una forma mucho más enmarañada y dispersa entre los párrafos y las tablas del manuscrito, y no merecen tantos comentarios como los beneficios en el apartado de la discusión. A pesar de las normas CONSORT[12], que no todos los editores aplican y exigen, esto es una realidad, y ni siquiera tales normas han desarrollado todavía un enfoque detallado del problema. La necesidad de hacer un buen balance riesgo-beneficio ha de estar siempre presente ante una toma de decisión, y en estudios oncológicos o sobre técnicas quirúrgicas mucho más que en otros, porque no es infrecuente que ciertos avances se produzcan a costa de mayor toxicidad o de otro tipo de problemas para los enfermos.

El lector, digamos, "de tipo medio" ha de tomar entonces ciertas cautelas que, normalmente, consistirán en hacer por su cuenta aquello que, sin mala fe, no suelen hacer los autores: evaluar más detalladamente el balance riesgo-beneficio. En nuestra opinión, la forma más simple, eficaz y directa de hacerlo (existen complejísimos métodos para llevarlo a cabo) es utilizar dos medidas de la importancia clínica de un resultado: el número de pacientes que es necesario tratar (NNT) y el número de intervenciones necesario para causar daño (NND)[12]. NNT es beneficio y NND se refiere al riesgo. Si la diferencia entre la probabilidad de supervivencia de dos tratamientos tras cierto tiempo de seguimiento es del 20%, NNT = 1 / 0,2 = 5. Se lee como que hay que tratar a 5 pacientes con el nuevo tratamiento para observar un beneficio adicional en supervivencia, es decir, que no se observaría con el tratamiento estándar. Este número de enfermos representa el promerio en esfuerzo terapéutico a realizar para obtener tal beneficio adicional. El NND se calcula igual pero con respecto a las complicaciones que sean, eso sí, importantes para el paciente. Siempre pueden extraerse de una o más tablas o de párrafos sueltos del texto, sumarlas para cada grupo y calcular el inverso de su diferencia en tanto por uno.

Si además de ser más beneficioso que el que constituye el grupo control, el nuevo tratamiento es menos dañino, el balance es obvio y no hace falta cuantificarlo para la toma de decisión. Pero si, como ocurre muy a menudo, los daños del nuevo tratamiento son superiores a los del tratamiento clásico, es necesario establecer el cociente entre NND y NNT colocando siempre en el numerador el NND y el NNT en el denominador[13,14]. Lo apropiado es poner un ejemplo real y, para ello, nada mejor que echar mano de la comparación entre tratamientos neoadyuvantes o perioperatorios frente a cirugía sola en el tratamiento de ciertos tumores muy frecuentes, como los de mama, esófago-cardias o recto, muy en boga en los últimos tiempos y con tendencia a aumentar. La neoadyuvancia tiene sus propias complicaciones que se suman y en ocasiones interaccionan con las quirúrgicas, por lo tanto, siempre las habrá en mayor número que con la cirugía sola. Cuantificar el balance se hace pues imprescindible.

Dado que han sido apuntados por algunos autores[15], recurriremos a los resultados (NNT y NND) "promedio" o aproximados que se desprenden de diversos ensayos clínicos muy conocidos[16-18], en los que se comparan tratamientos perioperatorios o neoadyuvantes frente a cirugía sola en el cáncer de esófago y cardias. El NND suele estar alrededor de 20 en contra de la combinación de quimioterapia y cirugía, mientras que hasta la fecha el NNT suele estar en 10 a favor de dicha combinación. Si 20 / 10 = 2, se traduce como que podemos esperar 1 daño adicional por cada 2 pacientes en los que se haya logrado beneficio. Y ahora viene ya la toma de decisión fundada: ¿esto es mucho o poco rendimiento? Lo conveniente entonces es dibujar lo que nos ocurriría a nosotros mismos si, en nuestra casuística, estas cifras se dieran de esta forma. Por ejemplo y como es nuestro caso, que tratamos anualmente 10-12 casos de este tumor con posibilidades de resección, en principio, curativa. A la sazón, podemos imaginar que, si aplicáramos a todos ellos el tratamiento combinado, en el transcurso de 2 años se beneficiaría a unos 2 pacientes, se dañaría seriamente a 1 paciente y se perdería tiempo y recursos en alrededor de 20 casos a los que ni se beneficiaría ni se dañaría, y en los que se podría haber acudido directamente a la cirugía. El problema está en que no sabemos de antemano qué pacientes concretos serán los que caerán en un lado u otro, hasta que se descubra alguna forma segura de saberlo. Así, ante este balance tan discreto, seguramente lo mejor sería esperar a que estos tratamientos combinados en este tumor sean una

firme realidad más que una buena esperanza, que es lo que parecen indicar estas cifras por ahora. Otra decisión sensata sería seleccionar para quimioterapia perioperatoria los casos potencialmente más "peligrosos" por avanzados, fuera de lo que correspondería a un ensayo clínico y dentro de lo que es la práctica diaria habitual, hasta que los conocimientos sobre la cuestión se refinen y amplíen. Resulta evidente que realizar un balance riesgo-beneficio un poco detallado puede cambiar la percepción de un resultado que se desprende de la lectura cruda de un artículo, y casi siempre actúa como moderador de optimismos exagerados.

Bibliografía

1. Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in clinical trials. J Chron Dis. 1967;20:637-48.
2. Man-Son-Hing M, Laupacis A, O'Rourke K, Molnar FJ, Mahon J, Chan K, et al. Determination of the clinical importance of study results. A review. J Gen Intern Med. 2002;17:469-76.
3. Chan K, Man-Son-Hing M, Molnar FJ, Laupacis A. How well is the clinical importance of study results reported. An assessment of randomized controlled trials. CMAJ. 2001;165:1197-202.
4. Jaeschke R, Singer J, Guyatt GH. Measurement of health status: ascertaining the minimal clinically important difference. Control Clin Trials. 1989;10:407-15.
5. Van Walraven C, Mahon JL, Moher D, Bohn C, Laupacis A. Surveying physicians to determine the minimal important difference: implications for sample size calculations. J Clin Epidemiol. 1999;52: 717-23.
6. Prieto-Valiente L, Herranz-Tejedor I. ¿Qué significa "estadísticamente significativo"? La falacia del criterio del 5% en la investigación científica. Madrid: Díaz de Santos; 2005.
7. Escrig-Sos J, Miralles-Tena JM, Martínez-Ramos D, Rivadulla-Serrano I. Intervalos de confianza: por qué usarlos. Cir Esp. 2007; 81:121-5.
8. Katz MH. Multivariable analysis. A practical guide for clinicians. Cambridge: Cambridge University Press; 2000.
9. Mills JL. Data torturing. N Engl J Med. 1993;329:1196-9.
10. Bender R, Lange S. Adjusting for multiple testing. When and how. J Clin Epidemiol. 2001;54:343-9.
11. Perneger TV. What's wrong with Bonferroni adjustments. BMJ. 1998;316:1236-8.
12. Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, et al. Improving the quality of reporting of rabdomizaed controlled trials. The CONSORT statement. JAMA. 1996;276:637-9.
13. Glasziou P, Guyatt GH, Gordon H, Dans AL, Dans LF, Strauss S, et al. Applying the results of trials and systematic reviews to individual patients. Evidence-based medicine. 1998;3:165-6.
14. Simon SD. Statistical evidence in medicals trials. What do the data really tell us? New York: Oxford University Press; 2006.
15. Gee DN, Rattner DW. Management of gastroesophageal tumors. Oncologist. 2007;12:175-85.
16. Kaklamanos IG, Walker GR, Ferry K, Franceschi D, Livingstone AS. Neoadjuvant treatment for resectable cancer of the esophagus and the gastroesophageal junction: a meta-analysis of randomized clinical trials. Ann Surg Oncol. 2003;10:754-61.
17. Fiorica F, Di Bona D, Schepis F, Licata A, Shahied L, Venturi A, et al. Preoperative chemoradiotherapy for oesophageal cancer: a systematic review and meta-analysis. Gut. 2004;53:925-30.
18. Cunningham D, Allum WH, Stenning SP, Thompson JN, Van de Velde CJ, Nicolson M, et al. Perioperative chemotherapy versus surgery alone for resectable gastroesophageal cancer. N Engl J Med. 2006;355:11-20.