

# Ronda clínica y epidemiológica

OSCAR OSÍO URIBE<sup>1</sup>, JOHN JAIRO ZULETA TOBÓN<sup>2</sup>

## ANÁLISIS DEL DESENLACE EN ESTUDIOS EXPERIMENTALES

**E**L PRESENTE ARTÍCULO DE LA RONDA CLÍNICA y Epidemiológica está orientado a dar algunas pautas a los lectores acerca de cómo deben interpretar los artículos de las revistas médicas con desenlaces (endpoints) múltiples, esto es, cuando los investigadores miden la asociación entre una exposición y varios resultados, unos establecidos previamente como “primarios” y otros como “secundarios” o cuando cada uno de estos resultados está conformado por grupos de variables. Si tanto en los desenlaces primarios como en los secundarios, y en algunos casos hasta en los terciarios, se relata una asociación similar con la exposición, no hay muchas dificultades en la interpretación del fenómeno. Los problemas de interpretación surgen cuando en el curso de un estudio clínico experimental la exposición va seguida de unos desenlaces divergentes o contrarios, en los cuales se nos informa que se encontraron resultados “positivos” en los desenlaces secundarios (es decir, que hubo una asociación con significancia estadística), pero que esta situación no se dio en los desenlaces primarios, o también, cuando se escribe que la asociación se dio en los resultados primarios, pero que no tuvo significación estadística. Veamos entonces cuáles son algunos de los problemas que tienen los clínicos cuando escogen las variables de desenlace en el curso de un estudio clínico experimental, aspectos que incidirán en la viabilidad del proyecto, en el análisis posterior de los resultados y en la determinación que harán los lectores de la validez del experimento.

1. Escogencia de un resultado sustituto, adaptado, de aquello que realmente se desea evaluar, con el fin de reducir el tamaño de la muestra. Ejemplo: escoger

.....  
<sup>1</sup> Especialista en Medicina Interna, Mg Sc en Epidemiología Clínica, Director de la Corporación Académica de Patologías Tropicales, Universidad de Antioquia, sajoro@epm.net.co  
<sup>2</sup> Profesor de Ginecología y Obstetricia, MgSc Epidemiología clínica. Universidad de Antioquia.

como desenlace el efecto que tiene un medicamento sobre la presión arterial en vez de medir la disminución que se presentará de los ataques cerebrovasculares.

2. Elección de una variable de desenlace que idealmente debe ser continua o de razón, para que pueda ser medida con exactitud y precisión. Las variables continuas tienen la ventaja sobre las dicotómicas (nominales) de aumentar la potencia del experimento y reducir hasta en un 50% el tamaño de la muestra necesaria para llevarlo a cabo. Si es necesario medir desenlaces nominales, la potencia dependerá más del número de acontecimientos o eventos que se produzcan durante el seguimiento que del tamaño de la muestra estudiada.
3. Respeto por el enmascaramiento o cegatura: la variable de desenlace debe ser medida por evaluadores que no conozcan a qué grupo pertenece el paciente (bien sea al de control o al de tratamiento activo).
4. Diseño de estrategias investigativas que limiten el número de sujetos que desertan por cualquier causa del estudio, luego de la aleatorización, y a quienes no se les puedan medir las variables de desenlace.
5. Selección de un número suficiente, pero no ilimitado, de variables de desenlace, porque al escoger un número plural se pueden medir diferentes aspectos del fenómeno que se está investigando, y al limitarlas a las necesarias se reduce el papel que puede jugar el azar en la interpretación de los resultados. Se recomienda que en el diseño de un experimento que evalúa un tratamiento, se piense en variables de desenlace que midan los efectos adversos que potencialmente pueda tener la intervención.

El análisis e interpretación de los desenlaces compuestos es complejo, porque en realidad

durante el transcurso del experimento no se está analizando una sola hipótesis, sino que se están examinando hipótesis múltiples, situación que origina que algunos estadísticos rigurosos sugieran la corrección del nivel de significación estadística. El método de Bonferroni consiste en dividir el nivel de significación estadística por ejemplo 0.05, entre el número de hipótesis que se pretende probar (si, por ejemplo, hubiera cuatro hipótesis, se probaría cada una de ellas con un alfa de 0.0125, que resulta de dividir 0.05 entre cuatro). La estrategia de formular hipótesis múltiples no relacionadas o independientes tiene ventajas como son aumentar la eficiencia del estudio, permitiendo responder a más preguntas con un solo experimento, descubrir una mayor cantidad de asociaciones, aumentar el poder del tiempo necesario para el análisis estadístico (sólo cuando las asociaciones en los distintos desenlaces se dan en la misma dirección) y evitándole al clínico escoger un solo desenlace. Cuando las hipótesis están relacionadas entre sí, al encontrarse asociaciones consistentes, las conclusiones serán más firmes. Una regla de oro en los estudios clínicos es el establecimiento previo de todas las hipótesis que el investigador quiera hacer, pero definir una de ellas como la hipótesis primaria, porque esto ayuda a que el estudio se centre en su objetivo principal y proporciona una base clara para el cálculo del tamaño de la muestra. Además, cuando se tiene un desenlace primario identificado previamente, se puede demostrar la significación estadística de la hipótesis principal, sin tener que apelar a la corrección sugerida cuando se tienen hipótesis múltiples. Las asociaciones no esperadas, así sean estadísticamente significativas, aparecidas durante el análisis de los datos son fuente de ideas interesantes que le permiten al investigador explorar o producir nuevas hipótesis y experimentos, pero que mientras tanto deben considerarse como asociaciones potenciales que tienen una gran probabilidad de haber sido producidas por el azar.

En la revista British Medical Journal de Octubre del año 2005, en la sección "Educación y Debate", el señor Nick Freemantle, quien es un estadístico que ha participado en varios estudios clínicos de diabetes, criticó que durante el último Congreso de la Asociación Europea para el Estudio de la Diabetes se hubieran presentado los resultados preliminares de un estudio prospectivo que evaluó el efecto de la pioglitazona en la enfermedad aterosclerótica macrovascular, amañando y malinterpretando los resultados. Explicó su comentario diciendo que los autores del estudio habían faltado a la verdad al sostener que el antidiabético oral había logrado demostrar, por primera vez para esta familia de medicamentos, ser capaz de reducir un desenlace compuesto, integrado por la mortalidad por todas las causas, el infarto del miocardio no fatal y los ataques cerebrovasculares. El señor Freemantle señaló que esa conclusión no estaba apoyada en una práctica epidemiológica sólida ni respetaba los estándares de análisis de los resultados de este tipo de estudios. Veamos por qué.

El estudio PROACTIVE, que es el trabajo en mención, suministró pioglitazona o placebo de manera aleatoria a 5.000 diabéticos de alto riesgo inadecuadamente controlados, muchos de los cuales venían recibiendo tratamiento para distintas formas de enfermedad cardiovascular. El desenlace primario fue definido como un compuesto integrado por la mortalidad por todas las causas, el infarto del miocardio no fatal, el accidente cerebrovascular, la amputación de miembro inferior, el síndrome coronario agudo y las revascularizaciones; los investigadores informaron que la asociación había tenido una  $p$  "grande" de 0.10 (superior a 0.05) o sea, estadísticamente no significativa. El desenlace definido como secundario era también un compuesto de resultados, y los investigadores informaron que la asociación tuvo una  $p$  "pequeña", de 0.03 (significativa); este fue el desenlace mostrado como exitoso para el beneficio proporcionado por el

medicamento. La presentación criticada del PROACTIVE hizo énfasis entonces en el beneficio (estadísticamente significativo) que el fármaco había demostrado en algunos desenlaces secundarios, pero dejó de lado que en su desenlace primario el beneficio no había alcanzado esa misma significación. Dice el señor Freemantle que esta presentación dejó de lado un hecho aceptado por la mayoría de los estadísticos: el desenlace primario no necesariamente es el resultado clínicamente más importante, pero sí es el fundamental desde el punto de vista estadístico, porque es el que da la pauta principal para la interpretación del estudio. Cuando el desenlace primario no muestra una asociación significativa, todo el "alfa" o error tipo uno se "gasta" en la búsqueda de la asociación y no queda nada para la demostración de los desenlaces secundarios; el autor de la crítica concluye que el desenlace secundario de un estudio clínico sólo tiene un significado "nominal". La significación estadística en un desenlace secundario de un estudio clínico aleatorizado, cuando es contrario a la asociación encontrada del desenlace primario, sugiere una asociación para explorar, no para considerarla como una hipótesis demostrada.

Como conclusión de las razones de su crítica, el señor Freemantle puntualiza que como en el PROACTIVE el medicamento antidiabético demostró que en los desenlaces primarios compuestos no era significativa su acción, una afirmación acerca de sus beneficios basada en sus desenlaces secundarios compuestos no tiene la suficiente fortaleza estadística para probar una asociación; sugirió que un juicio definitivo se aplase para después del debate académico que se dé una vez que la investigación se publique en una revista médica.

## BIBLIOGRAFÍA

1. HULLEY S, FEIGAL D, MARTIN M, CUMMINGS S. Diseño de un nuevo estudio: Experimentos. En: Diseño de la Investigación Clínica, 1ª ed.

Barcelona: Ediciones Doyma, SA; 1993: 123-140.

2. BROWNER WS, NEWMAN TB, CUMMINGS S, HULLEY SB. Preparación para estimar el tamaño de la muestra: hipótesis y principios subyacentes. En: Diseño de la Investigación Clínica. Barcelona: Ediciones Doyma, SA; 1993: 141-151.
3. FREEMANTLE N. How well does the evidence on pioglitazone back up researchers' claims for a reduction in macrovascular events. *BMJ* 2005; 331: 836-838.

**TÍTULO DEL ARTÍCULO:** Effect of treatment of gestational diabetes mellitus on pregnancy outcomes.

**AUTORES:** CROWTHER C, HILLER J, MOSS J, MCPHEE A, JEFFRIES W, ROBINSON J, et al. para el Grupo Australiano de estudio de la intolerancia a los carbohidratos en las mujeres embarazadas (ARCHOIS)

**REFERENCIA:** *New Engl J Med* 2005; 24 (352): 2.477- 2.486.

**DISEÑO:** experimento clínico en el cual se distribuyeron al azar mil diabéticas gestacionales con embarazos entre 24 y 34 semanas, para recibir tratamiento con dieta, control glicémico e insulina en caso necesario (490) o recibir el cuidado prenatal rutinario sin informarles su diagnóstico y sin hacerles ninguna intervención para controlar la diabetes. Las características clínicas y demográficas de los grupos fueron similares al inicio del estudio y no hubo pérdidas del seguimiento. Como resultado principal del estudio se evaluó un desenlace compuesto llamado "complicaciones perinatales severas" que integra en uno solo la muerte perinatal, la distocia de hombros, la

fractura ósea y la lesión nerviosa. Como resultado secundario se analizaron el ingreso a la unidad de cuidado neonatal y la ictericia. El desenlace para evaluar a la madre igualmente fue integrado y tuvo en cuenta la necesidad de realizar cesárea o de hacer inducción del parto.

**RESULTADOS:** hubo una diferencia estadísticamente significativa a favor de la intervención en el desenlace primario (1 frente a 4%) con un riesgo relativo de 0.33 (IC<sub>95%</sub> de 0.14-0.75). En el grupo con manejo de la diabetes se presentaron más casos de ingreso del recién nacido a la unidad especial (RR 1.13 IC<sub>95%</sub> 1.03-1.23) y hubo más casos de inducción del trabajo de parto (RR 1.36, IC<sub>95%</sub> 1.15-1.62). No hubo diferencias en la proporción de cesáreas. Para los componentes del desenlace primario se encontró: 0.3% distocias de hombros en el grupo de intervención frente a 1.4% en el grupo de control, lo que se traduce en un RR de 0.45 (IC<sub>95%</sub> 0.19-1.09), fracturas 0.2% frente a ninguna, muertes perinatales 5 en el grupo de intervención y 1 en el grupo control (RR 0.5 IC<sub>95%</sub> 0.2-1.0) y parálisis nerviosas 0.6 frente a 0%, respectivamente.

**ANÁLISIS:** es un trabajo que cumple en buena medida con el rigor que requiere un experimento: hubo aleatorización confiable, enmascaramiento de la asignación, las pacientes estaban enmascaradas al diagnóstico, la información la recolectó un auxiliar de investigación aunque no hubo enmascaramiento de los médicos tratantes, se compararon dos grupos bastante similares, el seguimiento fue completo y se había planeado un análisis por intención de tratar que no se requirió; por lo tanto, se puede concluir que se trata de un estudio válido que amerita evaluar los resultados. Al momento de intentar aplicar estos resultados se encuentra que el tratamiento fue efectivo para disminuir en el recién nacido el riesgo de tener una complicación importante; sin embargo, fueron varias complicaciones medidas como una sola y al evaluarlas en forma

individual, la situación ya no parece tan clara. Esto lleva a resolver una serie de preguntas con el fin de aclarar la validez y la aplicabilidad del agrupamiento de desenlaces en uno solo.

El punto de partida para este análisis es identificar si todos los desenlaces que componen el desenlace compuesto tienen una importancia clínica similar. Es claro que en este estudio existe un desequilibrio importante entre los desenlaces seleccionados para la agrupación: una cosa es la distocia de hombros y otra bien diferente son las potenciales consecuencias graves de la misma. La mayoría de los casos de distocia de hombros se resuelven con maniobras obstétricas sin generar morbilidad importante al recién nacido y obviamente no tiene la misma importancia que la muerte. Igualmente, las lesiones nerviosas y las fracturas no se asemejan a la muerte. En segundo lugar, es necesario identificar si el efecto de la intervención sobre cada uno de los componentes del desenlace compuesto es similar y ante todo descartar que la frecuencia del desenlace menos relevante sea mucho mayor que la del clínicamente más importante. En este caso, la frecuencia de los eventos fue similar, pero la distocia de hombros contribuyó con un 69.6% de los desenlaces adversos en el grupo control (16 de los 23) mientras que la muerte explicó el 21.3%. En tercer lugar, es necesario identificar las posibles diferencias entre las reducciones relativas del riesgo y la precisión de sus intervalos de confianza; en este estudio se encuentra que la distocia de hombros explica la mitad de la reducción del riesgo absoluto visto en el grupo de intervención.

Cuando se presenta esta situación -los desenlaces tienen diferentes importancias clínicas, el desenlace menos importante es el que más aporta a la diferencia global entre los tratamientos, existen diferencias importantes entre los riesgos relativos- es obligatorio evaluar individualmente los desenlaces que hacen parte del compuesto. Al evaluar en forma independiente la presencia de distocia de hombros, la lesión nerviosa, la fractura

y la muerte se aprecia que la diferencia entre los grupos no es estadísticamente significativa, producto de la falta de poder al reducir la muestra. Es importante, de todas maneras, tener en cuenta que el grupo de intervención no tuvo muertes y el de control sí tuvo cinco, pero se podría concluir que la pregunta con respecto a la utilidad del tratamiento de las pacientes con diabetes gestacional aún sigue sin respuesta objetiva.

**TÍTULO DEL ARTÍCULO:** Broad spectrum antibiotics for preterm, prelabour rupture of fetal membranes: the ORACLE I randomized trial.

**AUTORES:** KENYON S, TAYLOR D, TARNOW-MORDI W, para el Grupo colaborativo ORACLE

**REFERENCIA:** The Lancet 2001; 357: 979-988.

**DISEÑO:** experimento clínico en el cual en forma aleatoria se asignaron 1.197 mujeres a recibir eritromicina, 1.212 a recibir amoxicilina-clavulónico, 1.192 a recibir simultáneamente ambos antibióticos y 1.225 a recibir placebo con el fin de evaluar el efecto en embarazadas con ruptura prematura de membranas pretérmino. El desenlace primario fue un agregado de muerte neonatal, enfermedad pulmonar crónica entendida como dependencia de oxígeno después de las 36 semanas de vida posnatal o anomalías cerebrales mayores en la ecografía antes del alta. Tuvo 18 desenlaces secundarios adicionales.

**RESULTADOS:** 12.7% de las mujeres que recibieron eritromicina y 12.7% de las que recibieron placebo presentaron el desenlace primario ( $p = 0.08$ ); cuando se hicieron los análisis descartando los embarazos gemelares, estas cifras fueron 11.2 y 14.4%, respectivamente ( $p = 0.02$ ). La combinación clavulónico-amoxicilina no sólo no mostró ventajas sobre el placebo sino que aumentó el riesgo de enterocolitis necrosante.

**ANÁLISIS:** el interés de este trabajo en este momento es evaluar la validez del desenlace seleccionado. La respuesta a la primera pregunta es que posiblemente los tres desenlaces agregados tengan una importancia similar para las pacientes: la enfermedad pulmonar crónica y la muerte pueden estar en un nivel similar, sin embargo, para aceptar el tercer elemento, anormalidad mayor en la ecografía cerebral, sería conveniente que los autores aclararan el impacto clínico de estas anomalías con el fin de tener un juicio más objetivo, pero se puede aceptar que son tan importantes como los otros dos resultados. Con respecto a la frecuencia de los efectos adversos evaluados, se puede decir que son similares en el grupo de tratamiento con eritromicina, lo que refuerza el supuesto de que realmente el tratamiento sí tiene relación con los resultados: 5.5% niños con enfermedad pulmonar crónica, 4.2% con anormalidades cerebrales graves y 5.9% muertos (en el grupo placebo estas cifras son: 6.2, 5.0 y 6.7, respectivamente). Extrañamente, a pesar de ser un experimento clínico, los autores no presentan ningún riesgo relativo, y por consiguiente ningún intervalo de confianza, para valorar la precisión de las diferencias encontradas en el sinnúmero de comparaciones que realizaron y se limitaron a informar valores de p. Independiente de esto, se podría aceptar que el desenlace primario seleccionado fue adecuado y pertinente para valorar la intervención propuesta.

Es indispensable valorar la validez del estudio: esta investigación cumple adecuadamente con la asignación aleatoria de los tratamientos, el enmascaramiento de la secuencia de la asignación, el enmascaramiento

en el seguimiento, la similitud entre los grupos de comparación y el seguimiento completo de las pacientes. Aparentemente es una investigación válida; sin embargo, tiene dos defectos graves que difícilmente se pueden identificar a menos que se conozca su historia: el primero es que los autores no reconocen que este informe hace parte de una investigación mayor que tenía como objetivo evaluar el efecto del antibiótico en las pacientes con amenaza de parto prematuro en general y no de la ruptura prematura de membranas en particular y que decidieron dividir el informe al no encontrar diferencias en el grupo inicial; esto se refleja en que, a pesar del rigor que muestran en muchos aspectos metodológicos en el informe, nunca informan los supuestos para el cálculo del tamaño de la muestra. El segundo problema es que presentan unos resultados de análisis de subgrupos que no aclaran que se hubieran planeado en el protocolo, como es el caso de analizar únicamente los embarazos simples de mujeres con ruptura prematura de membranas que fue en los únicos en que encontraron algunas diferencias. Una confirmación de esto es la aparición de una diferencia estadísticamente significativa al excluir los embarazos gemelares, diferencia que no existía al evaluar la totalidad de la población estudiada. Los análisis de subgrupos son considerados exploratorios y no confirmatorios y por lo tanto ameritan confirmación con trabajos diseñados específicamente para ello. En este caso en particular las diferencias encontradas podrían ser simple producto del azar dado que se hicieron más de 100 comparaciones. Un último aspecto es tener en cuenta la diferencia encontrada en este análisis de subgrupo: 11.2 frente a 14.4. Muy posiblemente aún se requieran más estudios para resolver esta pregunta de investigación.

