

El análisis de resultados quirúrgicos y el laberinto de la confusión

Aunque no lo parezca a simple vista, y a pesar del gran desconocimiento que hay acerca de la materia entre bastantes profesionales, lo cierto es que no resulta nada fácil analizar correctamente, sin sesgos, los resultados de la actividad quirúrgica. La causa está en la gran cantidad de factores que influyen en ella de forma importante, aparte del grado de capacitación de un cirujano concreto, o de un departamento quirúrgico, incluso de un centro hospitalario en su conjunto.

La actuación de estos factores de confusión es dramática cuando el objetivo es comparar resultados de técnicas quirúrgicas, o de morbilidad entre cirujanos, departamentos u hospitales, con respecto a una técnica o al tratamiento de una enfermedad, aparte de lo que son análisis aleatorizados. Ésta es una situación muy habitual o incluso necesaria cuando no hay otra alternativa que plantear un estudio retrospectivo o una auditoría interna o externa. Estos factores de confusión, sobre los que hay que ejercer algún tipo de control, suelen relacionarse con factores pronósticos que los pacientes traen consigo de su casa, y sobre los que poco o nada se puede hacer por modificarlos. Si se compararan directamente, sin más, los resultados de un cirujano que operara casos avanzados de un tumor con los de otro que operara exclusivamente casos tempranos, sería lógico inferir que, a igual competencia profesional, el último obtendría mejores resultados de supervivencia, incluso de morbilidad postoperatoria. Aquí la comparación determinaría un resultado no sólo injusto, sino estrictamente falseado, debido al efecto de confusión del estadio de la neoplasia. Siempre que una variable actúa, como la del estadio tumoral, correlacionada con el resultado de una intervención de cualquier tipo, y a su vez también correlacionada con la variable que es objeto de la comparación, como en este ejemplo, puede dar lugar a confusión si su efecto no se ajusta entre las partes que representan el sujeto de comparación. En realidad, esto podría ocurrir igual en cualquier clase de estudio no aleatorizado, aunque no se refiera a resultados de una intervención de tipo quirúrgico.

El proceso de aleatorización no lleva a que los pacientes asignados a un grupo tengan un pronóstico similar, sino a que esta asignación no esté influida por los distintos grados de pronóstico que traen consigo los pacientes elegidos. Así, en su ausencia, hay que hacer algo activamente para imitar a posteriori, de alguna manera, los efectos de la aleatorización, aunque no sea absolutamente perfecta. Este algo consiste en utilizar técnicas de análisis estadístico de datos que ejerzan ese control o

ajuste de forma automatizada, como son las técnicas denominadas multivariantes. En el caso de los resultados quirúrgicos, en los que se suele tratar variables cualitativas binarias (mortalidad sí-no, complicaciones sí-no, etc.), la regresión logística (RL) será la que obtenga el protagonismo principal con mucha diferencia. En lo fundamental, esta técnica analiza un resultado binario en función de la variable principal que es objeto de la comparación. Pero si junto a ésta se introducen los factores de confusión que pudiera haber, el resultado obtenido con la variable que establece el objetivo a comparar quedará ajustado por el efecto de todos los factores que la acompañan en el modelo. Su uso puede hacerse de 2 modos: a) como la descrita, en que cada posible factor de confusión se introduce individualmente, o b) como paso preliminar que agrupe a estos factores en una escala de puntuación que con posterioridad se utilizará como una variable unificada, para introducirla en otro modelo final, junto con la variable a comparar. En el primer caso, la variable de resultado sería, por ejemplo, la mortalidad quirúrgica; en el segundo caso, y en el paso preliminar, la variable de respuesta podría ser el tipo de técnica quirúrgica, o el tipo de hospital o cirujano, y una vez obtenida la probabilidad de asignación según los factores de posible confusión, se confeccionaría el modelo final con la variable de resultado asistencial como respuesta, y la escala de puntuación antes obtenida junto con la variable a comparar (cirujano, hospital u otra) como predictoras. Las escalas de puntuación, como la POSSUM o la APACHE, son ejemplos de escalas ya prefabricadas que tienen sus indicaciones y sus inconvenientes para este uso. Pero según la cuestión a analizar, es posible que fuera más conveniente agrupar los factores de confusión en una sola variable creada para la ocasión. Ésta es una tecnología emergente, aunque no nueva, que se conoce como *propensity scores*¹, y que tiene su principal indicación en estudios no aleatorizados sobre resultados raros o infrecuentes en la comparación de técnicas comunes², como sería el caso del absceso intraperitoneal postoperatorio al comparar apendicectomía laparoscópica y convencional. La RL es el punto de partida del procedimiento.

Cuando el objeto a comparar ya no es el resultado de una técnica concreta frente a otra, sino los resultados de la asistencia quirúrgica, la base del procedimiento sigue siendo la misma, pero las cosas se complican mucho más. Aquí, la RL no es más que un elemento del proceso, imprescindible, pero no único. Entramos en lo que en lengua inglesa se conoce como *risk adjustment*³. Por lo

general, se basa en relacionar los episodios que se refieren a un resultado en cuanto al número observado y al número esperado, bien en forma de razón o ratio, bien en forma de diferencia entre ellos. El número de episodios esperados es la parte sobre la que influyen de forma directa todos los factores de confusión que pudieran intervenir, y en su obtención es donde, de nuevo, la RL es la protagonista, como norma general. Cualquier diferencia importante entre lo observado y lo esperado, sea en ratios o en diferencias propiamente dichas, suele ser el reflejo de diferencias en la calidad asistencial que tienen repercusión en el resultado concreto de esta asistencia. También aquí los investigadores se han preocupado de obtener escalas de puntuación que faciliten el análisis, pero su aplicación universal es más que dudosa, por lo que su uso debe estar precedido de un proceso de validación riguroso aplicado para una población concreta de pacientes y de agentes sanitarios. Otro punto crucial en el ajuste de riesgos, aparte del uso de la RL, es la forma de obtener los datos, ya que de ello depende que se recojan todos los factores de confusión implicados que pueda haber, y que se recojan de forma apropiada a la técnica de análisis a aplicar. De nuevo, aquí pueden encontrarse formularios prefabricados y bien adaptados a enfermedades concretas, como es el caso de la base de datos ASCOT para la cirugía del cáncer gastroesofágico^{4,5}. Con todo, el ajuste de riesgos debe ser un procedimiento específico de cada contexto. Así lo hace constar lezzoni⁶ en su libro, que es la biblia de esta metodología. Una estrategia apropiada de ajuste de riesgo necesitaría así responder a 4 preguntas básicas. La primera hace referencia a qué clase de resultado se refiere. La segunda ha de dirigirse a especificar el escenario temporal del análisis. La tercera ha de llevar a concretar el tipo de población de pacientes a la que se enfoca. La cuarta, el propósito sustantivo al que se encamina el estudio.

La repercusión sobre la población de pacientes y de médicos de la tecnología del ajuste de riesgos puede llegar a ser enorme. Recuérdese que es la base sobre la que se decide la centralización de enfermedades que se está llevando a cabo en ciertos países desarrollados. También es la base de los informes públicos dirigidos a los pacientes sobre resultados asistenciales de los diversos centros a los que pueden tener acceso. Esta ola aún no ha llegado a España de forma clara, pero hay indicios de que no tardará mucho en hacerlo, y no nos debería sorprender desprevénidos a los cirujanos, ni a los hospitales, para convertirnos en meros sujetos pasivos de lo que pueda ocurrir. Como se puede deducir de los párrafos anteriores, no es una cuestión fácil para nosotros. Seguramente, esta clase de análisis las realizarán técnicos ajenos a nuestra profesión y a nuestros centros, ya que requiere unos conocimientos de los que no disponemos los cirujanos, y las decisiones las tomarán los políticos de turno. ¿Qué podemos hacer ante ello? Tu et al⁷, en una revisión de conjunto, ofrecen un decálogo, del cual, personalmente, me atrevería a destacar 2 mandamientos sobre todos los demás. Ante todo, hay que conocer cómo funciona esta clase de análisis. No quiero decir que vayamos a dominar la RL con todo el resto del proceso, y todas sus modalidades, pero sí hay que saber qué hay y cómo funcionan en general, de la misma manera que sa-

bemos indicar un antibiótico, aunque no conozcamos su fórmula química. En segundo lugar, hemos de disponer de nuestros propios datos enfocados hacia la cuestión: hemos de conocer qué factores de confusión hay para cada grupo de enfermedades y sus resultados asistenciales, y recogerlos al alta de nuestros pacientes, igual que se recogen los habituales datos administrativos, demográficos y los diagnósticos principales. Para saber qué factores de confusión concurren, si no lo hemos estudiado particularmente, hay que recurrir a la bibliografía. Hay información sobrada para ello. Lo demás es cuestión de disciplina a la hora de rellenar los diversos formularios que puedan necesitarse cuando se hace el informe de alta y guardarlos debidamente en soporte informático. El papel y la concienciación de los jefes de servicio resultan fundamentales como punto de partida. Tampoco debe despreciarse la ayuda de nuestras asociaciones profesionales para unificar y estandarizar una buena recogida de estos datos y debería ponerse en marcha cuanto antes. En otros lugares del mundo desarrollado ya se está llevando a cabo desde hace años. Así, tenemos que saber defendernos de la falta de rigor, de la mala fe, o de la competencia de intereses, que pudieran concurrir en un análisis externo de nuestra actividad asistencial. Por otra parte, el hecho de conocer sin sesgos nuestros propios resultados es la mejor forma de empezar a reparar las desviaciones que pudiesen existir.

El último aspecto a considerar es la forma de publicar una hipotética comparación de resultados asistenciales entre cirujanos u hospitales. No hay más que 2. Una es hacer una especie de clasificación o *ranking* en donde alguien se lleva el primer premio, una vez conocidos los resultados debidamente ajustados. Otra⁵ consiste en establecer una línea promedio que representen los resultados conjuntos de todos los individuos comparados, con sus bandas de fluctuación que delimiten por dentro lo que es un resultado simplemente normal o aceptable, por fuera del cual estaría en un lado la zona de excelencia, y en el lado opuesto, la zona, digamos, de ignominia. Los resultados ajustados de cada cual, incluso ajustados al volumen de asistencia⁵, caerían en alguna de estas 3 zonas. El primer enfoque es propio de países donde la asistencia sanitaria especializada está sujeta a las leyes de mercado y a la competencia entre centros. La segunda opción es más propia de países europeos donde esta asistencia está mayoritariamente colectivizada. En cualquier caso, pienso que esta segunda opción es más ecuánime y resultaría más fácil de comprender y asimilar en nuestro país, sobre todo en centros que pudieran presentar unos buenos resultados, a pesar de que tengan un menor volumen de enfermedad, ya que haberlos, seguro que los hay.

Javier Escrig Sos

Servicio de Cirugía General y Digestiva. Hospital General de Castellón. Castellón. España.

Bibliografía

1. Adamant M, Guller U, Weber P, Oertli D. Propensity scores and the surgeon. Br J Surg. 2006;93:389-94.

2. Braitman LE, Rosenbaum PR. Rare outcomes, common treatments: Analytic strategies using Propensity Scores. *Ann Intern Med.* 2002;137:693-6.
3. Daley J, Henderson WG, Khuri SF. Risk-adjusted surgical outcomes. *Annu Rev Med.* 2001;52:275-87.
4. Cumming J, McCulloch P. ASCOT: a comprehensive clinical database for gastro-esophageal cancer surgery. *Eur J Surg Oncol.* 2001;27:709-13.
5. Tekkis PP, McCulloch P, Steger AC, Benjamin IS, Poloniecki JD. Mortality control charts for comparing performance of surgical units: validation study using hospital mortality data. *BMJ.* 2003;326:786-90.
6. Iezzoni LI. Risk adjustment for measuring health care outcomes. 3rd ed. Chicago: AcademyHealth; 2003.
7. Tu JV, Schull MJ, Ferris LE, Hux JE, Redelmeier DA. Problems for clinical judgement: 4. Surviving in the report card era. *CMAJ.* 2001;164:1709-12.