

# Las variables de resultado combinadas: ¿útiles o engañosas?

DRES. GAIETÀ PERMANYER MIRALDA, IGNACIO FERREIRA GONZÁLEZ

## INTRODUCCIÓN

En la literatura médica actual es especialmente frecuente el uso de variables de resultado combinadas (VRC; habitualmente llamadas en inglés *composite endpoints* o *combined endpoints*). Esta frecuencia es claramente mayor en los estudios de terapéutica cardiovascular, como lo ha puesto de manifiesto una revisión reciente en la que, en los ensayos clínicos publicados en nueve revistas de elevado impacto, 64% de los relativos a enfermedades cardiovasculares o ictus usaban VRC <sup>(1)</sup>. Ello justifica el interés que los cardiólogos deberíamos tener por esta técnica de estudio. Resulta llamativo que, a pesar de su difusión, los análisis de su fundamento, auténtico significado, interpretación y limitaciones escasean en la literatura. Es, pues, pertinente llevar a cabo una valoración de cuáles son sus ventajas y problemas. Esta valoración podrá ser de utilidad para una lectura crítica de la literatura y constituir una guía, quizá provisional, para la interpretación de los estudios que usan este tipo de variables.

## DEFINICIÓN DE VARIABLE DE RESULTADO COMBINADA

Recuérdese que las variables de resultado constituyen aquellas descripciones cuantitativas o categóricas de fenómenos asociados al resultado de la evolución de la enfermedad o al efecto del tratamiento, y se denomina especialmente así a la expresión de los resultados de un ensayo clínico. La figura 1 pretende esquematizar qué se entiende por VRC. Tradicionalmente, las variables de resultado pueden ser varias (por ejemplo, desarrollo de muerte, infarto de miocardio, accidente cerebrovascular o necesidad de revascularización), pero suelen estar analizadas por separado, independientemente unas de otras, comparando el número de eventos correspondientes a cada una de ellas en el grupo de tratamiento activo y en el

grupo control. En las VRC, tal como ilustra la figura, lo que se contabiliza son los pacientes en los que se desarrolla por primera vez cualquiera de los mencionados acontecimientos. La comparación ilustrada en la figura, que se presenta como estadísticamente significativa, se refiere al número de pacientes en cada grupo terapéutico que han experimentado cualquiera de esos sucesos clínicos y en los que se cumple la VRC.

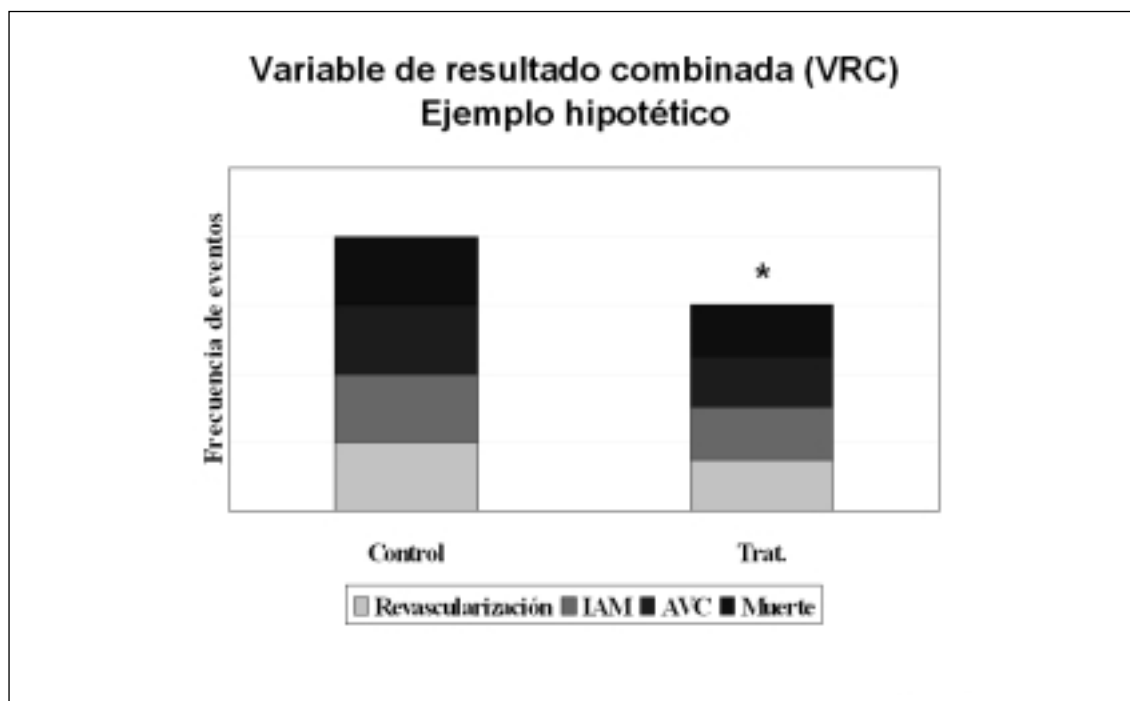
## PLANTEAMIENTO DEL PROBLEMA

A pesar de la simplicidad aparente del concepto de VRC, su comprensión e interpretación suscitan varios problemas sobre los que todavía no existe acuerdo ni resolución adecuados en la literatura médica. A continuación se expone la visión que del mismo tienen los autores. Para su mejor comprensión, es bueno comenzar recordando cómo se justificó durante la década de 1990 el uso de las VRC en ensayos clínicos. Las VRC se introdujeron fundamentalmente –concepto que no debe nunca perderse de vista– como una técnica para aumentar la eficiencia del análisis <sup>(2,3)</sup>. Se consideró que, al menos en medicina cardiovascular, el efecto de las intervenciones valoradas en la actualidad era casi sin excepción pequeño (al contrario que en ensayos clínicos de épocas previas), lo que determinaba que, si se deseaba valorar el efecto sobre la variable más importante, como es la mortalidad, la muestra o el período de seguimiento necesarios para alcanzar potencia estadística suficiente iban a ser inmensos, en el orden de varios miles de pacientes. La ventaja a causa de la cual se propuso la introducción de las VRC y por la que han alcanzado su gran difusión actual es la siguiente: si en lugar de considerar sólo una variable principal (digamos la mortalidad) se usa un conjunto

Unidad de Epidemiología, Servicio de Cardiología.

Hospital General Vall d'Hebron, Barcelona, España.

**Correspondencia:** Dr. G. Permanyer. Unidad de Epidemiología, Servicio de Cardiología. Hospital General Vall d'Hebron, Barcelona, España. Pº Vall d'Hebron 119-129. 08035 Barcelona, España



**FIGURA 1.** Representación hipotética de una variable de resultado combinada, constituida por la presentación, en el grupo control y en el grupo tratado, de muerte, ictus (AVC), infarto de miocardio (IAM) o necesidad de revascularización. El asterisco indica que la diferencia en la incidencia acumulativa de la VRC entre ambos grupos es estadísticamente significativa.

de acontecimientos desfavorables (por ejemplo, muerte, reinfarto, reingreso hospitalario, angina refractaria, etcétera) de tal manera que la aparición de cualquiera de ellos represente el cumplimiento de la variable, entonces el volumen de muestra requerido (y con éste el costo y duración del estudio) será menor. Inicialmente se consideró, por parte de algunos autores, que dichos acontecimientos añadidos a la variable principal deberían representar equivalentes (*surrogates*) de mal pronóstico, y también se sentaron varias normas para la construcción de VRC válidas <sup>(3)</sup>. No se ha llegado a acuerdo sobre ninguno de ambos puntos.

En el análisis del significado clínico e interpretación de las VRC nos planteamos dos preguntas fundamentales: 1) ¿En qué medida se aplica el efecto observado sobre la VRC a cada uno de sus componentes?, y 2) ¿Cuál es el significado clínico de la variable combinada como tal? En un plano más teórico, la primera pregunta podría formularse como: ¿Existen problemas de heterogeneidad en los componentes de la VRC?, y la segunda: ¿Cuál es la validez teórica de la VRC? En el texto que sigue se aborda cada uno de estos problemas por separado.

#### ¿EN QUÉ MEDIDA SE APLICA EL EFECTO OBSERVADO SOBRE LA VRC A CADA UNO DE SUS COMPONENTES?

La interpretación clínica de una VRC se mueve entre dos extremos. Supongamos que la VRC tiene cuatro componentes A, B, C y D. Un extremo representa entender que si el ensayo clínico demuestra un efecto sobre la VRC, puede suponerse que dicho efecto se ejerce sobre cada una de las variables componentes A, B, C y D. A esto puede llamarsele “interpretación cándida”. Se opone a ella la que puede llamarse “interpretación conservadora”, que consiste en, a la vista del efecto observado, limitarse a aceptar que puede afectar a alguna o algunas de sus variables componentes (A, B, C o D) sin que se pueda precisar cuál. Evidentemente, entre ambos extremos hay interpretaciones intermedias (por ejemplo, suponer que lo único que resulta plausible es el efecto sobre una de las variables componentes, no pronunciándose sobre el que pueda haber sobre las restantes).

#### FUNDAMENTO TEÓRICO DE ESTA INTERPRETACIÓN

Para abordar apropiadamente este problema es necesario referirnos brevemente, primero, al fundamento de teoría estadística que permitirá formular luego recomendaciones prácticas. La

razón más común para el uso de variables combinadas es el volumen excesivamente grande de la muestra necesaria para demostrar un solo efecto (mayor eficiencia). Ello equivale, por tanto, a la renuncia a evaluar ese efecto aisladamente a favor de evaluar una combinación de efectos. ¿Qué coherencia lógica tiene, pues, preguntarse *a posteriori* cuál ha sido ese efecto? Puede, sin embargo, objetarse que, aun si se acepta esta aparente incongruencia, puede ser legítimo especular sobre el efecto en las variables componentes. Si realmente pudieran sacarse conclusiones del efecto sobre éstas, realmente el empleo de las VRC sería un ardid maravilloso, ya que permitiría obtener conclusiones sobre algunas variables con eficiencia mucho mayor que considerándolas por separado.

Pero más allá de disquisiciones lógicas, existe una noción de teoría estadística que hace problemático deducir el efecto sobre las variables componentes de una a partir de la hipótesis sobre la que se ha formulado el ensayo clínico, a menos que se tomen una serie de precauciones (que usualmente no se tienen en cuenta) <sup>(4)</sup>.

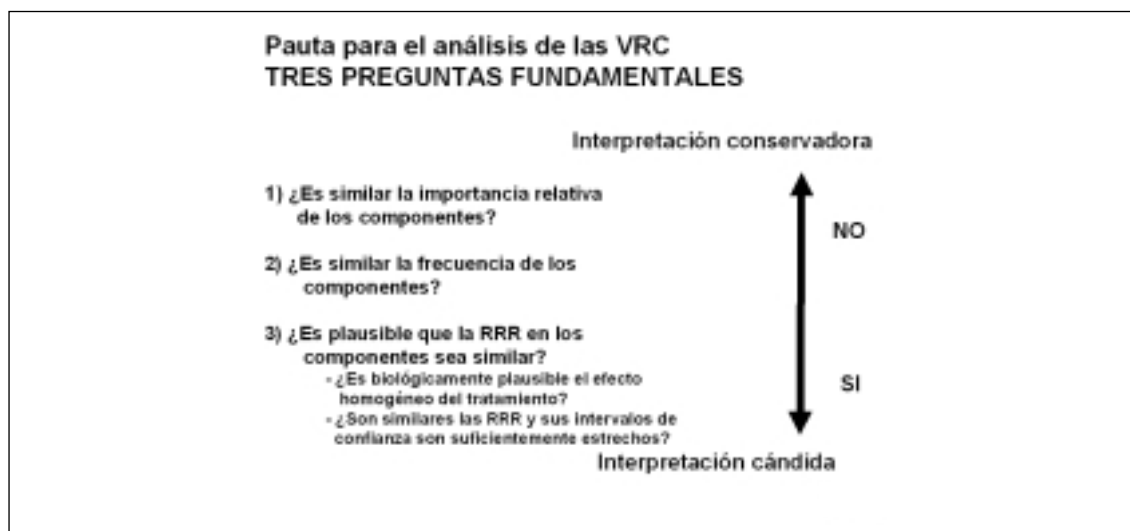
Un ejemplo permitirá comprender esto: siguiendo a Moyé <sup>(4)</sup>, imaginemos un ensayo clínico en el que los investigadores han planteado su hipótesis sobre la incidencia de una VRC consistente en muerte u hospitalización. Han calculado la muestra del ensayo basándose en aceptar un error tipo I de 0,05 para el análisis de dicha VRC. Al analizar sus resultados comprueban que el valor de *p* hallado para la diferencia de efecto entre los dos grupos de tratamiento es 0,09. O sea, que no existe diferencia significativa entre ambos tratamientos sobre la combinación de muerte y hospitalización. Pero se preguntan entonces sobre el efecto en los componentes de la VRC. Cuando analizan el efecto en los datos del estudio sobre la hospitalización y la muerte por separado, hallan un valor de *p* de 0,15 para la hospitalización y de 0,03 para la muerte. ¿Sería legítimo que los autores concluyeran, con sólo estos datos, que el efecto de un tratamiento sobre la muerte es significativo? No podría en absoluto llegarse a esta conclusión sobre la base de la hipótesis del estudio y el procedimiento usado para verificarla: la muestra del estudio se calculó aceptando una tasa de error tipo I de 0,05 para la VRC y sólo para ella. Cualquier otra comparación estadística expone al cálculo al riesgo de significación espuria debida a comparaciones múltiples.

Para evitar este riesgo, los autores hubieran debido distribuir el error tipo I entre los componentes de la VRC, lo que hubiera representado la necesidad de una muestra mucho mayor y, por tanto, una contradicción con el objetivo habitual de las VRC. Por tanto, no pueden valorarse las comparaciones estadísticas en los componentes de una VRC a menos que se tenga en cuenta la posibilidad de error tipo I. Ello equivale a exigir que los autores hayan formulado hipótesis sobre los componentes y ajustado apropiadamente el volumen de muestra o que el valor de *p* observado sea suficientemente pequeño para considerar improbable dicho error, lo que en la práctica es sumamente infrecuente.

Pero en realidad, aun admitiendo una cierta incoherencia lógica y las dificultades del contraste de hipótesis sobre los componentes de una VRC, el clínico puede todavía formular preguntas legítimas sobre éstos, si acepta el carácter *post hoc* y la naturaleza estimativa de las mismas. Es decir, existe un abordaje pragmático del problema que consiste en que una vez aceptado que la hipótesis contrastada en el ensayo clínico se refiere exclusivamente a la VRC y no a sus componentes, pueden analizarse *a posteriori* los efectos observados sobre los mismos y aplicar nociones de estimación estadística que permitan juzgar sobre la plausibilidad del efecto de la intervención sobre ellos.

#### ABORDAJE PRAGMÁTICO

No sólo existen algunas características del efecto observado en los componentes de la VRC que permiten afirmar, en algunos casos, que es plausible que exista un efecto sobre alguno de ellos, sino que en ocasiones se podrá incluso plantear la hipótesis de que dicho efecto se aplica de manera consistente a todos los componentes (*interpretación cándida*). Nosotros, en colaboración con el Departamento de Epidemiología de la Universidad Mc Master en Ontario <sup>(5)</sup>, hemos desarrollado un algoritmo (figura 2) para el análisis estimativo del efecto de una intervención sobre los componentes de la VRC según el cual, y de acuerdo a la respuesta que demos a tres preguntas relativas a la homogeneidad de dichos componentes y las características del efecto de la intervención, podremos estimar como más o menos plausible la existencia de un efecto homogéneo de la intervención sean sobre ellos. Cuanto más afirmativas las respuestas a dichas preguntas, más plausible será la *interpretación*



**FIGURA 2.** Esquema de un posible enfoque práctico para plantear la plausibilidad de que el efecto de una VRC esté (interpretación cándida) o no (interpretación conservadora) homogéneamente presente en los componentes de la misma. La primera interpretación será menos plausible cuanto menos respuestas sean afirmativas. Tomado de ref. 5.

*cándida* (efecto de la intervención sobre todos los componentes de la VRC), mientras que cuanto más negativas sean, más *conservadora* tendrá que ser nuestra interpretación (efecto sobre la VRC en su conjunto sin poder afirmar cuál es el efecto en los componentes).

Las tres preguntas son, como consta en la figura: 1) ¿Es similar la importancia relativa de los componentes? 2) ¿Es similar la frecuencia de los componentes? 3) ¿Es plausible que la reducción relativa del riesgo (RRR) en los componentes sea similar? Es decir: ¿Es biológicamente plausible el efecto homogéneo del tratamiento? ¿Son similares las RRR y sus intervalos de confianza son suficientemente homogéneos y estrechos? Es llamativo que los datos que aportan los ensayos clínicos no siempre permiten una respuesta a estas preguntas. No es excepcional, incluso en revistas muy divulgadas, que algunos investigadores comuniquen tan sólo el efecto sobre la VRC y no en todos sus componentes, o los intervalos de confianza del efecto sobre éstos. La comunicación incompleta de los resultados es uno de los malos usos de las VRC que debería corregirse.

*¿Es similar la importancia relativa de los componentes?*

Para decidir este punto intervendrá siempre un elemento de apreciación personal, ya que no es plausible que una taxonomía cualquiera aporte una gradación objetiva de importancia clínica. Sin ir más lejos, algunos podrían argumentar que siempre que en una VRC, como es

usual, figure la muerte, ello introducirá un gradiente de importancia con respecto a las otras variables. Pero esta apreciación puede ser irrelevante para los objetivos del estudio y, por otra parte, debe siempre distinguirse cuándo se propone una VRC con acontecimientos clínicos de importancia homogénea (incluyendo o no la muerte u otros acontecimientos con grave repercusión sobre la salud) de cuándo hay una disparidad manifiesta.

Considérese, por ejemplo, un ensayo clínico hipotético en el que se compare el efecto de dos técnicas percutáneas de revascularización miocárdica (tabla 1) y, para valorar el resultado, se emplee una VRC consistente en la aparición, a lo largo del seguimiento, de muerte, infarto de miocardio o necesidad de nueva revascularización del vaso culpable. Aquí no hay duda de que la importancia clínica de los componentes es claramente heterogénea, como lo sería en un ensayo que utilizara como VRC el infarto de miocardio, la necesidad de revascularización o la aparición de prueba de isquemia positiva. En general, la inclusión de acontecimientos basados, aun parcialmente, en alguna decisión médica (*physician-driven*), como es la necesidad de intervención o de reingreso hospitalario, o simplemente en la aparición de una prueba de laboratorio positiva, representan un elemento de heterogeneidad importante ya que en su conjunto (sea cual fuere la relevancia en casos aislados) su importancia clínica puede considerarse menor que la de la muerte o de acontecimientos clínicos con ma-

**TABLA 1.** COMPARACIÓN HIPOTÉTICA ENTRE DOS MODALIDADES DE REVASCULARIZACIÓN MIOCÁRDICA PERCUTÁNEA. SEGUIMIENTO A 12 MESES

	<i>Modalidad A</i>	<i>Modalidad B</i>	<i>Odds ratio</i> <i>(IC 95)</i>
Muerte	6%	6%	(0,15-4,81)
Infarto de miocardio	3%	4%	(0,50-6,43)
Necesidad de nueva revascularización	27%	65%	(0,12-0,49)
Variable de resultado combinada (muerte, infarto o nueva revascularización)	33%	67%	(0,15-0,46)

nifistas secuelas o consecuencias. En patología cardiovascular son frecuentes las VRC integradas por muerte, infarto de miocardio o ictus. Ello no sería forzosamente equivalente a heterogeneidad marcada, pero para asegurar una aceptable homogeneidad de la VRC debería exigirse (y no es norma habitual, estando la cuestión pendiente de mayor acuerdo), que se incluyeran criterios de gravedad o secuelas clínicas suficientes.

*¿Es similar la frecuencia de los componentes?*

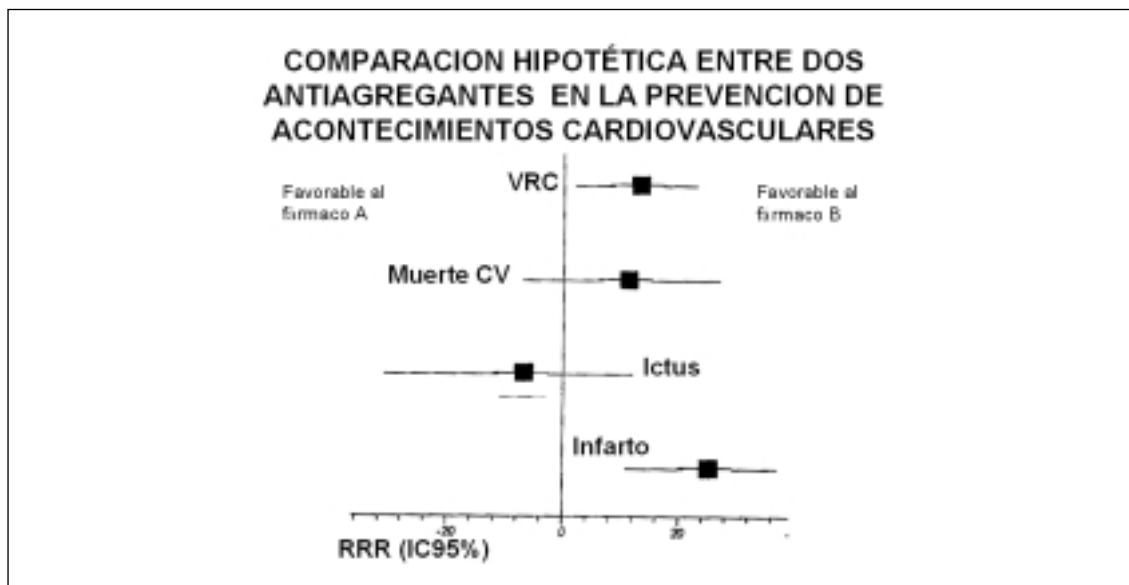
El segundo criterio de plausibilidad de un efecto homogéneo de la intervención estudiada sobre todos los componentes de la VRC es que estos no hayan acaecido con frecuencia marcadamente dispar. Una vez más, como es constante en este análisis, la interpretación de estos hechos no es inequívoca. Sin embargo, existen ejemplos marcados y frecuentes de la disparidad aludida. En el ejemplo de la tabla 1 se observa una tasa de muerte de 6% en ambas y de infarto de miocardio de 3% en una, y 4% en otra, mientras que la necesidad de revascularización es de 27% en la primera y de 65% en la segunda, lo que representa una diferencia superior a siete veces entre los acontecimientos más y menos frecuentes en cada grupo terapéutico. Aunque los autores presentaran sus resultados como una reducción significativa del efecto sobre la VRC, sería muy poco plausible suponer, dada la frecuencia tan heterogénea de acontecimientos, que el efecto estudiado se ejerce homogéneamente sobre los componentes. Es especialmente frecuente que en los estudios con gran disparidad en la frecuencia de los componentes la mayor frecuencia corresponda, como en el ejemplo hipotético expuesto, a la variable de menor relevancia clínica en su conjunto (en este caso la revascularización). En estos casos, e incluso en otros no tan inequívocos, es habitual que la comunica-

ción de los resultados, o su reseña posterior, se haga en un lenguaje ambiguo que favorece la sobrevaloración del efecto por parte del lector.

*¿Es plausible que la RRR en los componentes sea similar?*

También aquí el criterio del lector o investigador primará sobre las reglas fijas. De las dos cuestiones que integran esta pregunta, dejaremos de lado la plausibilidad biológica del efecto homogéneo del tratamiento, ya que parece muy difícil afirmar la implausibilidad salvo en casos extremos. Es más habitual tener que cuestionar si son similares las RRR y si los intervalos de confianza de las mismas son suficientemente homogéneos y estrechos. A este respecto tampoco hay normas fijas pero sí ejemplos ostensibles de falta de cumplimiento, sin hablar de aquellos casos ya mencionados en los que los autores no aportan datos suficientes para este examen. Volviendo al ejemplo de la tabla 1, se observan unos intervalos de confianza de la *odds ratio* entre 0,15 y 4,81 para la muerte, de 0,50 a 6,43 para el infarto (o sea, ambos llamativamente anchos y haciendo poco plausible el resultado en un sentido u otro), mientras que el intervalo de confianza de la *odds ratio* para nueva revascularización es mucho más estrecho (0,12-0,49) y, además, claramente compatible con un efecto de una técnica superior al de la otra.

En el ejemplo expuesto en la tabla, por tanto, la respuesta a las tres preguntas es negativa. En absoluto puede considerarse plausible un efecto homogéneo sobre los tres componentes de la VRC, es decir, una interpretación “cándida” no parece justificada. Según el análisis estadístico y la hipótesis del estudio (suponemos que no se hizo ajuste para comparaciones múltiples) sólo podemos afirmar que hay un efecto probable sobre la VRC, mientras que según nuestra estimación podemos suponer



**FIGURA 3.** Representación gráfica de la reducción relativa de riesgo en la comparación hipotética del efecto de dos antiagregantes plaquetarios en un ensayo clínico. Tanto los puntos correspondientes a las estaciones puntuales del efecto como correspondientes a sus intervalos de confianza (líneas) representan un efecto favorable al fármaco A cuando están situadas a la izquierda de la línea vertical (efecto 0) y favorable al fármaco B cuando están situados a su derecha.

como plausible que este efecto se ejerce sobre la variable más débil sin poder afirmar nada del efecto sobre los otros componentes.

No siempre las cosas son tan sencillas, y a veces esta interpretación pueda resultar opinable. Imaginemos un ensayo clínico en el que se comparan dos fármacos antiagregantes para la prevención de acontecimientos cardiovasculares, utilizando una VRC integrada por muerte, ictus e infarto de miocardio durante el seguimiento. Los resultados de ese ensayo hipotético se representan en el diagrama de la figura 3, en el que, como es usual en este tipo de gráfico, la línea vertical representa una RRR de 0 (ausencia de efecto), indicando tanto la estimación puntual como los intervalos de confianza a la derecha o izquierda de esta línea un efecto predominante de uno u otro fármaco. Como puede comprobarse, la RRR de la VRC es claramente favorable a uno de los fármacos (llamémosle B, con todo su intervalo de confianza a la derecha de la línea de ausencia de efecto). Si nos formulamos ahora las tres preguntas propuestas podemos llegar a las respuestas siguientes: 1) No existe un marcado gradiente de importancia clínica entre los componentes de la VRC (damos por supuesto que los autores han definido bien los ictus con secuelas y los infartos de miocardio no mínimos). 2) La frecuencia de aparición de acontecimientos no fue claramente dispar (este dato no se representa en

la figura). 3) Es más problemática la conclusión sobre una RRR concordante y homogénea. Los valores correspondientes al infarto de miocardio son similares, incluso más favorables al mismo fármaco. La estimación puntual de la RRR de muerte va en el mismo sentido, pero su intervalo de confianza, además de más ancho, abarca también la zona favorable al fármaco B. Su interpretación, por tanto, sería más dudosa. Pero el efecto sobre el ictus es llamativo: no sólo la estimación puntual del efecto del fármaco favorece al fármaco A, sino que la mayor parte del intervalo de confianza de la RRR es también favorable a él (en ausencia de normas concretas, parece razonable exigir que no haya discrepancias cualitativas o direccionales entre el efecto sobre los componentes, como las hay en este caso, para asumir un efecto homogéneo sobre los mismos). Concluimos, pues, que aunque la respuesta a las dos primeras preguntas es afirmativa, el comportamiento de las RRR no permite, en este ejemplo hipotético, suponer como muy plausible que el efecto de uno de los fármacos sea superior al otro de forma homogénea en la prevención de muerte, ictus e infarto, aunque el efecto sobre la combinación de acontecimientos lo sea.

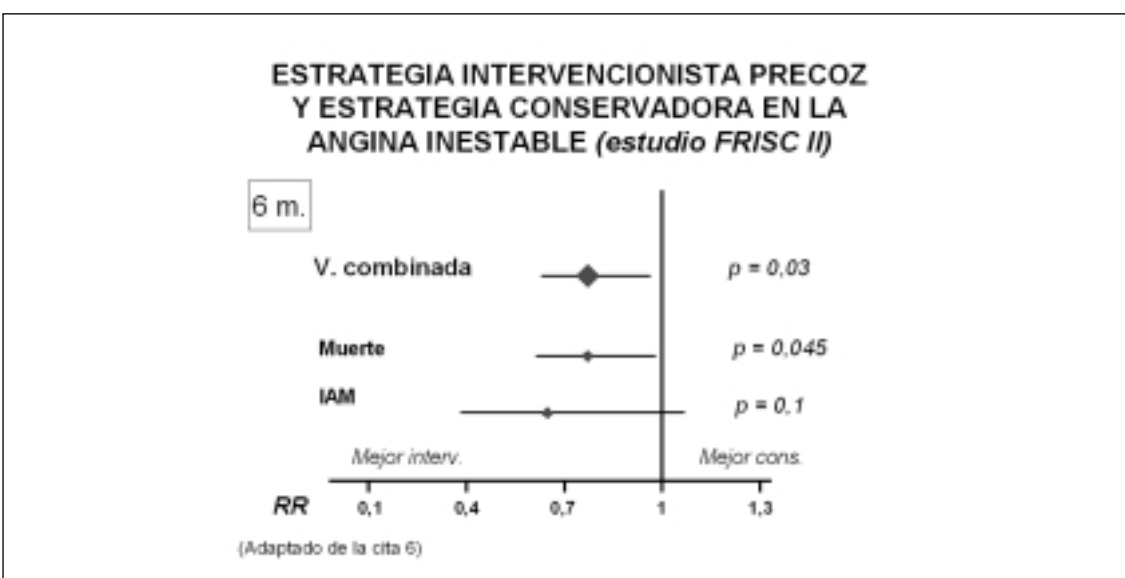
Lo habitual es que los ensayos clínicos con VRC no permitan una interpretación "cándida". No obstante, hay excepciones. Por ejemplo, el es-

**TABLA 2.** ESTRATEGIA INTERVENCIONISTA PRECOZ Y ESTRATEGIA CONSERVADORA EN LA ANGINA INESTABLE (ESTUDIO FRISC II)

6 m	Interv.	Cons.	RR (IC 95%)	p
VRC	113 (9,4%)	148 (12,1%)	0,78 (0,62 - 0,98)	0,031
Muerte	94 (7,8%)	124 (10,1%)	0,77 (0,60 - 0,99)	0,045
IAM	23 (1,9%)	36 (2,9%)	0,65 (0,39 - 1,08)	0,10
1 a				
VRC	127 (10,4%)	174 (14,1%)	0,74 (0,60 - 0,92)	0,005
Muerte	105 (8,6%)	143 (11,6%)	0,74 (0,59 - 0,94)	0,015
IAM	27 (2,2%)	48 (3,9%)	0,57 (0,30 - 0,90)	0,016

(Adaptado de las citas 6 y 7)

Interv: estrategia intervencionista precoz; Cons.: estrategia conservadora; RR: riesgo relativo; IC 95%: intervalo de confianza del 95%; IAM: infarto agudo de miocardio; VR: variable de resultado combinada (muerte o infarto de miocardio).



**FIGURA 4.** Representación esquemática con el mismo formato de la figura 3 de los resultados a seis meses del estudio FRISC II. Se representan los efectos sobre los riesgos relativos (RR) de muerte, infarto de miocardio (IAM) o cualquiera de ambos (V. combinada). Los resultados a la izquierda de la línea vertical son favorables a la estrategia intervencionista, y los situados a su derecha lo son a la estrategia conservadora.

tudio FRISC II <sup>(6)</sup> comparó una estrategia intervencionista precoz con una estrategia conservadora en el tratamiento de la angina inestable y usó una VRC de muerte e infarto de miocardio. A los seis meses de seguimiento el efecto sobre la VRC (tabla 2) era significativamente favorable a la intervención precoz, siéndolo sólo marginalmente para la muerte y en absoluto para el infarto de miocardio (no se había hecho, además, ajuste para comparaciones múltiples). Si nos formulamos ahora las preguntas de rigor, vemos que 1) Podemos aceptar que entre los dos componentes de la

VRC, según las definiciones del estudio, no había un excesivo gradiente de importancia. 2) Aunque la tasa de infarto fue inferior a la de muerte, la diferencia no era tan acusada como en casos en los que hemos considerado que la respuesta a esta pregunta debía ser negativa. 3) Las reducciones del riesgo relativo (relacionado con la RRR) eran bastante homogéneas (figura 4), con intervalos de confianza similares y compatibles con un efecto favorable a la intervención precoz. En este contexto, esto se aplica también al infarto de miocardio a pesar de su falta de significación estadística nominal

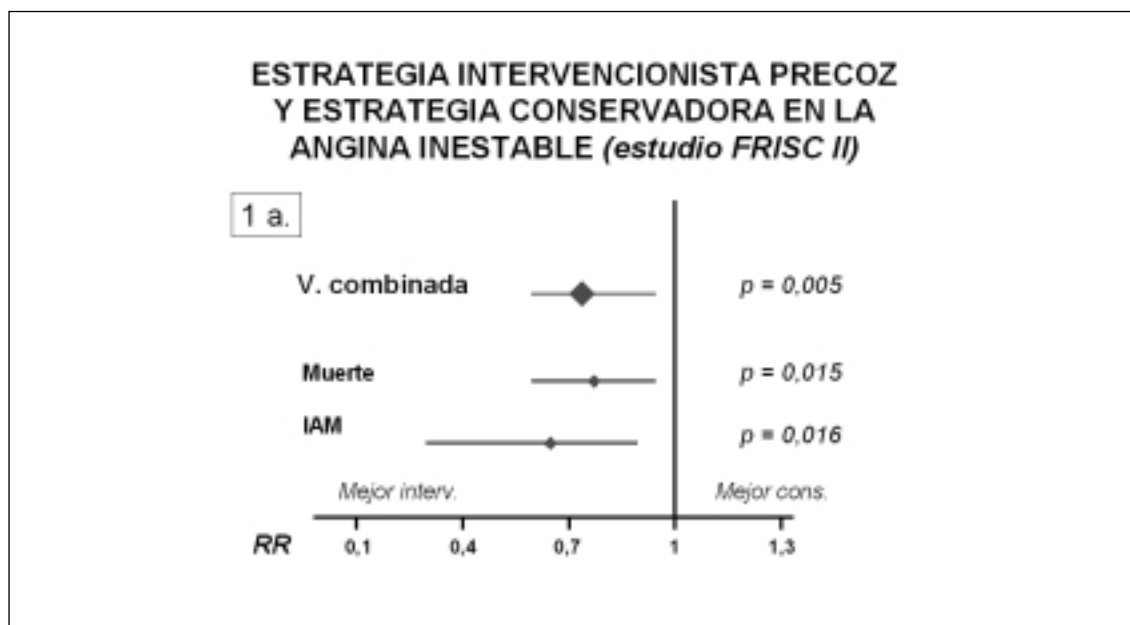


FIGURA 5. Resultados del estudio FRISC II al año de seguimiento. Formato igual al de la figura 4.

ya que su intervalo de confianza sólo franquea mínimamente la línea de ausencia de efecto. Podemos, por tanto, según estos resultados, considerar plausible que el efecto beneficioso de la intervención precoz sea aplicable tanto a la VRC como a la muerte y al infarto por separado.

Lo interesante de este ejemplo es que los autores publicaron posteriormente los resultados observados al año del seguimiento<sup>(7)</sup>. El mayor número de acontecimientos permitió ya observar unos hallazgos (tabla 2, figura 5) que confirman incuestionablemente la interpretación "cándida" que hubiera podido defenderse a los seis meses de seguimiento. Los intervalos de confianza se alejaban aun más de la línea de ausencia de efecto y, a mayor abundamiento, el efecto era estadísticamente significativo para los dos componentes a un nivel difícilmente explicable como espurio.

A modo de resumen de lo dicho hasta aquí, podemos afirmar que hay tres interpretaciones posibles de las variables combinadas:

- 1) El efecto más plausible se limita a uno o unos pocos componentes de la VRC.
- 2) Hay razones para suponer que el efecto es más o menos homogéneo en todos los componentes (interpretación "cándida").
- 3) Sólo puede aceptarse un efecto sobre la VRC en conjunto. No se puede asumir ningún efecto sobre los componentes (interpretación conservadora).

¿CUÁL ES EL SIGNIFICADO BIOLÓGICO Y CLÍNICO DE LA VARIABLE COMBINADA COMO TAL?

Como se ha afirmado en la primera parte de este artículo, esta segunda pregunta, fundamental para la comprensión de las VRC, equivale a interrogarnos sobre la validez teórica de las mismas. Quede ya dicho desde este momento que este es un tema complejo e insuficientemente abordado en la literatura del que no va a ser posible hacer, a pesar de su trascendencia, un abordaje sistemático. Comentaremos, sin embargo, algunos aspectos importantes del mismo.

A pesar de la escasez de la literatura correspondiente, los significados o las funciones que se han atribuido a las VRC son variados<sup>(2,4,8-10)</sup>. Entre ellos figuran: el constituir una medida del efecto de la intervención sobre el impacto global de la enfermedad (conjunto de acontecimientos desfavorables)<sup>(2,8)</sup>; ser una medida del efecto de la intervención sobre la tasa de pacientes vivos y libres de acontecimientos<sup>(9)</sup>; ser una medida del efecto de la intervención sobre la mortalidad y sus variables sustitutorias (*subrogadas*)<sup>(3)</sup>, y constituir un recurso para la eliminación del efecto confusor de los riesgos competitivos<sup>(4,9,10)</sup>. Pero, con mucho, el uso más frecuente de las VRC es, sin más, en cuanto recurso de diseño y análisis para poner de manifiesto efectos pequeños de forma más eficiente, sin que los investigadores, en gene-



ral, sean explícitos sobre qué otra finalidad expresa les ha llevado al uso de las mismas.

El primero de los significados expuestos (la expresión del efecto sobre un conjunto de acontecimientos desfavorables) justifica nuestra atención en este momento al asociarse con lo que hemos llamado “interpretación conservadora” de las VRC. Según esta interpretación, lo único que podríamos deducir de la VRC es el efecto conjunto de la intervención estudiada sobre el conjunto de acontecimientos desfavorables que, en determinadas situaciones, representan el impacto de la enfermedad, sin que podamos pronunciarnos sobre el efecto preciso sobre cada uno de estos acontecimientos. Esta interpretación no significa, en absoluto, la invalidez de la VRC. Antes bien, podría decirse, por razones que hemos expuesto en párrafos anteriores, que corresponde a su significado inicial más estricto.

Pero una cosa es la validez y otra la relevancia o el significado clínico. ¿Cuándo o cómo podrá pensarse que una VRC de la que sólo sea posible una interpretación conservadora tendrá valor clínico? ¿Hay criterios o normas para establecerlo?

El problema de la relevancia clínica de un efecto demostrado como estadísticamente significativo es complejo y escapa del ámbito de esta discusión. Sin embargo, es evidente que uno de los criterios de dicha relevancia es el tamaño o la magnitud de ese efecto. Podemos, como ejemplo, contrastar dos ensayos clínicos reales que usan VRC de las que parece prudente establecer una interpretación conservadora.

En el primero de ellos <sup>(11)</sup> se comparó la eficacia de los diuréticos con la de los inhibidores de la enzima de conversión de la angiotensina (IECA) como primer tratamiento de la hipertensión en pacientes mayores. La variable de resultado principal era la combinación de muerte de cualquier causa y totalidad de acontecimientos cardiovasculares. Entre éstos se incluían no menos de 11 componentes, cuya gravedad oscilaba desde la muerte cardiovascular a un simple accidente isquémico transitorio cerebral. Los autores consideraron que los IECA demostraron una eficacia significativamente superior a los diuréticos. A la vista de la composición de la VRC y de los resultados concretos del estudio, resulta evidente que sólo era posible llegar a la interpretación conservadora; es decir, que el efecto se ejercía sobre el conjunto de la VRC sin poderse presumir efecto alguno sobre sus componentes (incluso con-

siderando que el efecto sobre el desarrollo de infarto de miocardio era significativo a un nivel de 0,05, ya que no parece haberse hecho ajuste alguno para las múltiples comparaciones). ¿Qué relevancia clínica tiene este hallazgo? Una razón para suponer que ésta no es grande es la magnitud del efecto: se observó una reducción relativa del riesgo de sufrir la VRC (cualquiera de sus múltiples componentes) de 11% (intervalo de confianza de 95%: 0-21) y una reducción absoluta de tan sólo 1,4%.

Podemos contrastar la magnitud del efecto sobre esta VRC con el observado en otro ensayo clínico reciente <sup>(12)</sup> en el que se comparaba, en 427 pacientes con infarto agudo de miocardio, la intervención coronaria percutánea de rescate tras trombólisis fallida con la nueva trombólisis y el tratamiento conservador en el infarto agudo de miocardio (estudio REACT). Se utilizó una VRC constituida por mortalidad total, mortalidad cardiovascular, nuevo infarto de miocardio, accidente cerebrovascular e insuficiencia cardíaca. El resultado del ensayo fue favorable a la intervención coronaria percutánea, con una tasa de la VRC de 15% frente a 30% y 31% en los otros dos grupos. El pequeño número de acontecimientos relativos a cada componente permitía suponer que los intervalos de confianza del efecto sobre éstos (no referidos, tampoco esta vez, en el artículo) serían suficientemente amplios como para no permitir aceptar la homogeneidad del efecto sobre ninguno de ellos. El efecto sobre el infarto de miocardio era significativo, pero, una vez más, su valor resulta limitado al no haberse ajustado según comparaciones múltiples. En este ensayo sólo parece razonable, por tanto, aceptar la interpretación conservadora del efecto sobre la VRC. Pero la magnitud del efecto es suficientemente grande (reducción absoluta del riesgo de 15%) como para considerar que el valor clínico de este resultado es notable aun aceptando la dificultad de suponer cuál ha sido el efecto sobre los componentes. Parafraseando una comunicación personal de un eminente epidemiólogo, podemos aceptar que la interpretación conservadora del efecto sobre una VRC será clínicamente valiosa cuando, aun aceptando que el efecto se limitara a la variable de menor relevancia clínica, su magnitud tendría todavía valor para el clínico o sería preferible para el paciente.

Penetramos así en un terreno complejo, debatible y, en todo caso, no privativo de las VRC: ¿cuál es la reducción absoluta del riesgo que el clínico debe considerar valiosa? La respuesta va a depender del impacto sobre la sa-

lud que tenga el acontecimiento considerado, de las preferencias del paciente, de la opinión del clínico y, en un último término, de los juicios valorativos de la sociedad. En realidad, en el debate sobre la importancia otorgada a cualquier acontecimiento clínico y a su prevención o curación participan muchas disciplinas y, en último término, su marco final lo constituyen los valores de una cultura o sociedad determinados.

Aunque estas últimas consideraciones se alejan ya de la problemática intrínseca de las VRC, no debe olvidarse que éstas fueron introducidas primordialmente para detectar eficientemente efectos terapéuticos de escasa magnitud. Es relevante para nuestro asunto, por tanto, juzgar en qué medida esa magnitud puede o debe influir en la toma de decisiones clínicas. De hecho, los trabajos que usan VRC representan un potencial para la sobreinterpretación del efecto en dos sentidos: en primer lugar, cuando, como ya se ha dicho, se refieren o comentan sus resultados en términos que dejan sobreentendido que el efecto sobre la VRC se aplica a todos sus componentes; y, en segundo lugar, porque la propia semántica de la exposición de lo que semejan múltiples efectos de un mismo fármaco puede conducir a una interpretación aumentada de la importancia clínica de los hallazgos del estudio.

Como conclusiones de lo expuesto podemos destacar las siguientes:

- 1) Las VRC son un recurso de diseño frecuentemente usado y poco analizado para aumentar la eficiencia de los ensayos clínicos.
- 2) En algunas ocasiones las características de los resultados permiten estimar el efecto sobre alguno o todos los componentes de la VRC.
- 3) Otras veces sólo es posible juzgar sobre la existencia de un efecto global sin poder estimar el efecto sobre los componentes.
- 4) Particularmente en este último caso el significado clínico de la VRC dependerá de la magnitud del efecto global y de la relevancia clínica que se atribuya a éste.
- 5) El juicio sobre dicha relevancia no depende sólo de consideraciones científicas sino que debe fundamentarse en los valores de los

pacientes, de los clínicos y de la comunidad en general.

- 6) Una interpretación superficial de las VRC facilita la sobrevaloración de la intervención, en especial cuando no se aportan datos suficientes sobre el efecto en los componentes.

#### BIBLIOGRAFÍA

1. **Freemantle N, Calvert M, Wood J, Eastaugh J, Griffin C.** Composite outcomes in randomized trials: greater precision but with greater uncertainty? *JAMA* 2003; 289: 2554-9.
2. **Braunwald E, Cannon CP, McCabe CH.** An approach to evaluating thrombolytic therapy in acute myocardial infarction. The 'unsatisfactory outcome' end point. *Circulation* 1992; 86: 683-7.
3. **Cannon CP.** Clinical perspectives on the use of composite endpoints. *Control Clin Trials* 1997; 18: 517-29.
4. **Moye LA.** Multiple analyses in clinical trials. Springer; 2003.
5. **Montori VM, Permyer-Miralda G, Ferreira-Gonzalez I, Busse JW, Pacheco-Huergo V, Bryant D, et al.** Validity of composite end points in clinical trials. *Br Med J* 2005; 330: 594-6.
6. **FRISC II investigators.** Invasive compared with non-invasive treatment in unstable coronary-artery disease: FRISC II prospective randomized multicentre study. *Lancet* 1999; 354:708-15.
7. **Wallentin L, Lagerqvist B, Husted S, Kontny F, Stahle E, Swahn E.** Outcome at 1 year after an invasive compared with a non invasive strategy disease: the FRISC II invasive randomised trial. *Lancet* 2000; 356: 9-16.
8. **Gent M.** Some issues in the construction and use of clusters of outcome events. *Control Clin Trials* 1996;18:546-9.
9. **Lubsen J, Kirwan BA.** Combined endpoints: can we use them? *Stat Med* 2002; 21: 2959-70.
10. **Hallstrom AP, Litwin PE, Weaver WD.** A method of assigning scores to the components of a composite outcome: an example from the MITI trial. *Control Clin Trials* 1992; 13: 148-55.
11. **Wing LM, Reid CM, Ryan P, Beilin LJ, Brown MA, Jennings GL, et al.** A comparison of outcomes with angiotensin-converting—enzyme inhibitors and diuretics for hypertension in the elderly. *N Engl J Med.* 2003;348:583-92.
12. **Gerslick AH, Stephens-Lloyd A, Hughes S, Abrams KR, Stevens SE, Ure NG.** Rescue angioplasty after failed thrombolytic therapy for acute myocardial infarction. *N Engl J Med* 2005; 353: 2758-68.