

Variables de resultado combinadas en los ensayos clínicos

Ignacio Ferreira-González^{a,b}, Pablo Alonso-Coello^{a,c}, Ivan Solà^{a,c}, Valeria Pacheco-Huergo^{a,d}, Antònia Domingo-Salvany^{a,e}, Jordi Alonso^{a,e}, Víctor Montori^{a,f} y Gaietà Permanyer-Miralda^{a,b}

^aCIBER de Epidemiología y Salud Pública (CIBERESP).

^bUnidad de Epidemiología. Servicio de Cardiología. Hospital Vall d'Hebron. Barcelona. España.

^cCentro Cochrane Iberoamericano. Servicio de Epidemiología Clínica y Salud Pública. Hospital de Sant Pau. Universidad Autónoma de Barcelona. Barcelona. España.

^dCentro de Atención Primaria Turó. Barcelona. España.

^eUnitat de Recerca en Serveis Sanitaris. IMIM-Hospital del Mar. Barcelona. España.

^fKnowledge and Encounter Research Unit. Department of Medicine. Mayo Clinic College of Medicine. Minnesota. Estados Unidos.

Las variables de resultado combinadas en los ensayos clínicos son un recurso metodológico usado con frecuencia, especialmente en los estudios cardiovasculares. Las motivaciones más importantes para su utilización son aumentar la potencia estadística del estudio, valorar el beneficio neto de una intervención y evitar una interpretación errónea del resultado en presencia de riesgos competitivos. Sin embargo, su interpretación puede ser problemática si hay heterogeneidad entre los componentes en cuanto a su importancia, la frecuencia de eventos o el efecto de la intervención. En la discusión que sigue se presenta un revisión conceptual de los problemas del uso y la interpretación de las variables de resultado combinadas en ensayos clínicos, especialmente los cardiovasculares. Se presenta además una sencilla guía de interpretación de los resultados de los estudios que utilizan variables de resultado combinadas a partir de la cual se puede valorar nuestra confianza en dichos resultados. Finalmente, se presenta un estudio empírico sobre cuál ha sido el uso real de variables de resultado combinadas potencialmente problemáticas en ensayos clínicos cardiovasculares.

Palabras clave: *Ensayos clínicos. Variables de resultado combinadas. Heterogeneidad.*

Composite Endpoints in Clinical Trials

Composite endpoints are often used in clinical trials, especially in the cardiovascular area. Decreases in sample size requirements, ability to assess the net effect of an intervention and to avoid bias in presence of competing risk are the most cited advantages for their use. However, there is a risk of misinterpretation when heterogeneity among components with respect to either importance, number of events or magnitude of treatment effect exist. In the following review we present a conceptual discussion about the rationale and interpretation of such variables. Also, a user's friendly guide to interpret the results of clinical trials based on composite endpoints is presented. We also present an empirical study that provides evidence of the use of misleading composite endpoints in cardiovascular clinical trials.

Key words: *Clinical trials. Composite endpoints. Heterogeneity.*

Full English text available from: www.revespcardiol.org

INTRODUCCIÓN

Uno de los retos más importantes en la valoración del efecto de una intervención terapéutica es la elección de la variable de resultado principal¹. Esta variable, que representa la hipótesis que dio lugar al estudio de investigación, debería ser clínicamente relevante (importante para el paciente), fácilmente evaluable, libre de sesgo, barata de determinar y sensible a la inter-

acción en estudio². Es difícil encontrar una única variable de resultado que presente todas las características mencionadas. Por otro lado, la valoración del efecto de las intervenciones es compleja, pues habitualmente una única intervención suele actuar en varios aspectos de un proceso fisiopatológico o bien que, actuando en uno solo, afecta a varios órganos y sistemas. Además, existe la posibilidad de que aparezcan efectos secundarios, tanto previsibles como desconocidos. Por todo ello, no es extraño que el investigador formule más de una hipótesis principal. Por ejemplo, el investigador podría formular las siguientes preguntas: ¿la intervención A es eficaz para reducir la mortalidad en la población en estudio?; ¿la intervención A es eficaz para reducir la proporción de infartos de mio-

Correspondencia: Dr. G. Permanyer-Miralda.
Unidad de Epidemiología. Servicio de Cardiología. Hospital Vall d'Hebron.
Pg. Vall d'Hebron, 119-129. 08035 Barcelona. España.
Correo electrónico: gpermany@gmail.com

ABREVIATURAS

ECA: ensayo clínico aleatorizado.
IAM: infarto agudo de miocardio.
RRR: reducción relativa del riesgo.
VRC: variable de resultado combinada.

cardio no fatales?; ¿la intervención A es eficaz para reducir el número de ictus no fatales?

La secuencia lógica para contestar a esas tres preguntas sería la realización de un ensayo clínico aleatorizado (ECA) con tres variables de resultado principales: muerte, infarto agudo de miocardio (IAM) no fatal e ictus no fatal. Una vez concluido, se presentarían los resultados de la intervención sobre las tres variables. Sin embargo, en la literatura médica actual son cada vez más frecuentes los ensayos clínicos en los que varias variables de resultado se reúnen en única variable de resultado, que se denomina variable de resultado combinada (VRC). Las VRC (*composite/combined endpoints* o *composite/combined outcomes*) engloban en un único estimador el número de pacientes que presentan al menos uno de los eventos de los componentes incluidos³. En el ejemplo anterior, la VRC correspondiente sería muerte o infarto de miocardio no mortal o ictus. Esta estrategia «sintetiza» las tres preguntas anteriores en una única: ¿la intervención A es eficaz para reducir la variable muerte o infarto no fatal o ictus no fatal? Claramente hay una diferencia en la naturaleza de ambos tipos de estrategias. Dicha diferencia radica en que en el caso de las VRC el análisis que a posteriori se realice con los componentes no corresponderá a la hipótesis principal del estudio.

En la discusión que sigue, basada en una revisión sistemática sobre el uso de las VRC principales⁴, se analizan críticamente cuáles son las motivaciones para su utilización, así como sus limitaciones y riesgos. Se presenta, además, una guía para su interpretación⁵. Finalmente, presentamos evidencia empírica del uso problemático de las VRC en ECA cardiovasculares⁶.

MOTIVACIONES PARA EL USO DE VARIABLES DE RESULTADO COMBINADAS

Una reciente revisión sistemática sobre la lógica y las limitaciones de las VRC localizó 17 artículos y un libro sobre este tema⁴. Esta revisión mostró que el conocimiento teórico de los problemas de las VRC está aún insuficientemente abordado y que los metodólogos discrepan sobre diversos aspectos relacionados con el uso de estas variables. La revisión identificó tres situaciones fundamentales en las que es habitual el uso de las VRC:

Disminución del tamaño de muestra requerido para demostrar efectos

Siguiendo el ejemplo anterior, supongamos que un investigador decide realizar un ECA para evaluar las tres hipótesis principales. Dicha estrategia conlleva el cálculo del tamaño muestral para contrastar cada una de ellas. En dicho cálculo intervienen los siguientes parámetros: la proporción esperada de eventos, los riesgos de un error tipo I (α) y tipo II (β), y la magnitud del efecto esperado de la intervención. Supongamos que el investigador quisiera demostrar una reducción relativa del riesgo (RRR) del 20% sobre cada una de las tres variables principales, con una potencia estadística del 80% (error tipo II del 20%) y un error de tipo I del 5%. El cálculo del tamaño de la muestra arroja las siguientes cifras: en el caso de la variable IAM, serían necesarios 1.000 pacientes; en el caso de la variable ictus, serían necesarios 5.000 pacientes; en el caso de la variable muerte, serían necesarios 20.000 pacientes. Si el investigador quisiera poner a prueba las 3 hipótesis anteriores con suficiente poder estadístico, debería incluir un mínimo de 20.000 pacientes.

Supongamos ahora que el investigador decide realizar un ECA utilizando la VRC «muerte o IAM no fatal o ictus no fatal». El efecto esperado de la intervención es, como en el caso anterior, del 20%. El error tipo I y el error tipo II son los mismos que en el caso anterior. Sin embargo, el cuarto parámetro, la proporción esperada de eventos, es claramente diferente: la proporción de pacientes que, a lo largo del seguimiento, sufrirán al menos uno de los eventos es sustancialmente superior. Este hecho se traduce en una disminución del tamaño de muestra, que será mayor cuanto menor sea el nivel de dependencia o correlación entre los componentes¹. De esta manera, en el «mejor» de los escenarios (en el caso que la correlación entre los tres eventos tomados dos a dos fuera negativa) con únicamente 1.000 pacientes se podría demostrar una RRR del 20% sobre la VRC.

Actualmente cada vez es más difícil demostrar el efecto de una intervención, ya que la mayoría de los pacientes están medicados y su pronóstico es considerablemente mejor que hace unas décadas. Esta circunstancia requiere la realización de ensayos con un tamaño de muestra cada vez mayor (*megatrials*), y ampliar el tiempo de seguimiento para conseguir un número suficiente de eventos (muertes, infartos, etc.). Todo ello complica la realización de este tipo de ensayos por motivos tanto logísticos como de coste. Una estrategia para superar esta circunstancia son las VRC.

Ahora bien, ¿qué significa que la intervención A reduce un 20% el RRR de la variable «muerte o IAM no fatal o ictus no fatal»?; ¿qué podemos inferir acerca del efecto de la intervención en cada uno de los componentes?

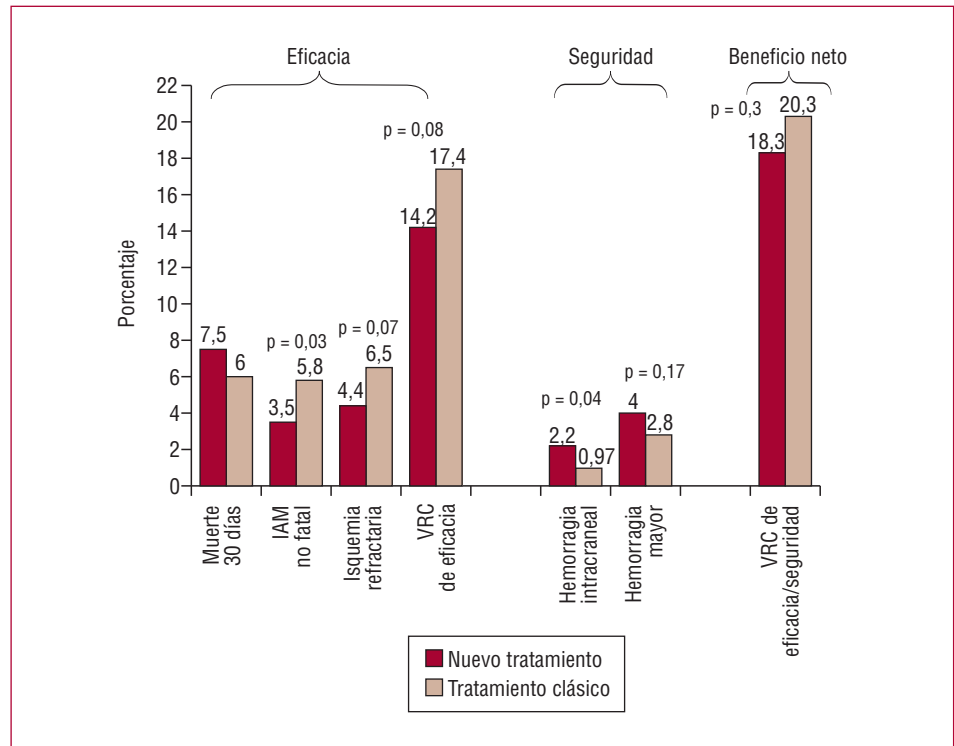


Fig. 1. Resultados de un ensayo clínico para valorar la eficacia de un nuevo régimen terapéutico asociado a tenecteplasa. IAM: infarto agudo de miocardio; VRC: variable de resultado combinada.

Valoración del efecto «neto» de una intervención

Supóngase ahora que la intervención en cuestión conlleva un riesgo asociado clínicamente relevante. Por ejemplo, se quiere valorar un nuevo trombolítico para el tratamiento del IAM. El tratamiento trombolítico supone un riesgo aumentado de hemorragia cerebral, y el investigador cree que hay razones para suponer que el nuevo trombolítico, mucho más eficaz que el estándar, conlleva un mayor riesgo. Se decide realizar un ECA con ambos tipos de trombolítico considerando la VRC «muerte o hemorragia cerebral».

En este escenario, la motivación para la utilización de una VRC no puede ser la disminución del tamaño de la muestra, puesto que es esperable que la dirección del efecto de la intervención en la variable muerte sea opuesta a la dirección del efecto en la variable hemorragia cerebral (es esperable que el trombolítico sea eficaz para reducir el número de muertes, pero que aumente el número de hemorragias cerebrales respecto al tratamiento estándar). La utilización de dicha VRC va a ser menos eficiente para demostrar un efecto que lo que sería, por ejemplo, la VR aislada «muerte», con la que es esperable una mayor reducción neta de eventos. ¿Cuál es, pues, la motivación para considerar dicha VRC? Simplemente la de capturar el «beneficio neto de la intervención». En el ejemplo anterior, de poco nos serviría un nuevo trombolítico que redujera algo la mortalidad pero aumentara de forma mucho

más sustancial el número de hemorragias cerebrales. Una estrategia sencilla para valorar el efecto de intervenciones que conllevan riesgos clínicamente relevantes es la utilización de una VRC que combine variables de «eficacia» y «seguridad». Si la nueva intervención reduce de forma estadísticamente significativa el porcentaje de eventos de la VRC, podemos confiar en que dicha intervención es en general más beneficiosa que la estándar.

Un ejemplo parecido se muestra en la figura 1, en relación a un EC reciente⁷ que valoró la eficacia y la seguridad de tenecteplasa junto con heparina de bajo peso molecular frente al mismo trombolítico asociado con heparina no fraccionada. Aunque la nueva combinación fue más beneficiosa que la clásica en variables de resultado de «eficacia», no ocurrió lo mismo con variables de resultado de «seguridad», en las que el nuevo tratamiento resultó deletéreo. De esta manera en la VRC global, que expresa el beneficio neto global del nuevo tratamiento, se observó una clara reducción de dicho beneficio.

Ahora bien, supóngase que el investigador tiene razones para creer que el nuevo trombolítico es eficaz para reducir el tamaño del IAM y decide considerar la VRC «muerte o hemorragia cerebral o presencia de nuevas ondas Q patológicas en el electrocardiograma». Los resultados muestran un beneficio sustancial asociado al nuevo tratamiento a expensas de una reducción del número de pacientes con «nuevas ondas Q patológicas», pero con un aumento del número de

hemorragias cerebrales y escaso efecto en la mortalidad. ¿Podemos confiar de la misma manera en esta VRC para capturar el beneficio neto de la intervención? ¿Se podría catalogar el nuevo trombolítico como generalmente más beneficioso que el estándar en el caso de observar una reducción estadísticamente significativa en dicha VRC? En este caso resulta problemático «confiar» en el resultado global de la intervención, ya que hay un gradiente de importancia muy acusado entre los componentes de la VRC. La variable «presencia de nuevas ondas Q patológicas», con una menor relevancia clínica, está influyendo en el resultado final con un peso similar a las otras dos variables (muerte y hemorragia cerebral) y consigue desplazar el efecto neto hacia un potencial beneficio sesgado.

Valoración del efecto en presencia de riesgos competitivos

En otras ocasiones, la razón fundamental para el uso de una VRC no es la posible reducción del tamaño de muestra ni la necesidad de capturar el beneficio neto de una intervención, sino evitar un sesgo en la valoración de un efecto en presencia de riesgos competitivos.

La posibilidad de un sesgo por riesgos competitivos tiene lugar en situaciones en que cuando se produce un evento disminuye la probabilidad de que se produzca otro evento de interés. Por ejemplo, supongamos que el investigador tiene razones para suponer que una intervención disminuye el riesgo de IAM no fatal. Los pacientes que fallecen antes de sufrir el evento de interés tienen un riesgo de sufrir IAM no fatal de cero. En este caso el «evento competitivo» es la muerte. Imagínese que un tratamiento no tiene ningún efecto en el riesgo de IAM pero que durante el ensayo clínico, de forma aleatoria o por efectos secundarios asociados al nuevo tratamiento, se producen más muertes en el grupo del nuevo tratamiento que en el grupo de pacientes asignados al control. El «riesgo» global de IAM en el grupo tratamiento es menor, pues hay un menor número de pacientes-años de seguimiento disponibles en el grupo tratamiento. De esta manera, si comparamos la tasa de IAM entre ambos grupos, podría ser que el tratamiento pareciera falsamente eficaz para disminuir el número de infartos de miocardio. Ahora bien, si en vez de utilizar la variable de resultado individual «IAM no fatal» se utiliza la VRC «muerte o IAM no fatal», el posible sesgo por riesgos competitivos queda anulado, pues ambas variables son analíticamente equivalentes.

Una situación parecida puede observarse en un ECA reciente que analizó la eficacia de los fibratos en prevención primaria cardiovascular en pacientes diabéticos, en el cual el IAM no fatal fue una de las variables de resultado de interés⁸. La incidencia de infartos no fatales en el grupo tratamiento fue significativamente

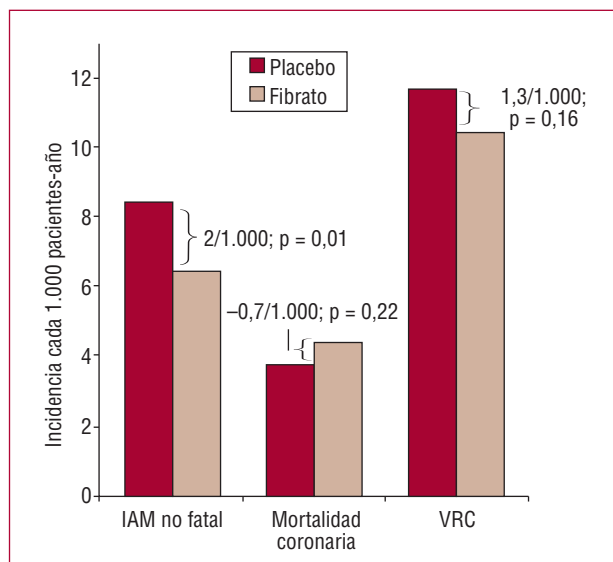


Fig. 2. Resultados de un ensayo clínico aleatorizado de fibratos frente a placebo en prevención cardiovascular primaria. IAM: infarto agudo de miocardio; VRC: variable de resultado combinada.

menor que en el grupo placebo: 6,4/1.000 pacientes-año en riesgo frente a 8,4/1.000 pacientes-año en riesgo ($p = 0,01$). Sin embargo, la incidencia de muerte de origen coronario fue algo mayor en el grupo tratamiento: 4,4/1.000 pacientes-año frente a 3,7/1.000 pacientes-año ($p = 0,22$) (fig. 2). El análisis de la VRC «muerte de origen coronario o IAM no fatal» no mostró un beneficio estadísticamente significativo: 10,4/1.000 en el grupo tratamiento frente a 11,7/1.000 en el grupo control ($p = 0,16$). De esta manera, no podemos descartar la posibilidad de que el efecto del tratamiento observado en la variable «IAM no fatal» sea parcialmente irreal debido a la presencia de riesgos competitivos. En este ejemplo concreto, la ausencia de efecto observado en la VRC ilustra cómo ésta puede ser útil para evitar un posible efecto espurio observado en uno de los componentes de la variable a causa de un sesgo por riesgos competitivos.

Por último, hay que mencionar que el análisis de los riesgos competitivos no es en absoluto competencia exclusiva de los ensayos clínicos aleatorizados. De hecho, dicha estrategia de análisis ha sido aplicada en estudios de cohortes, por ejemplo en el ámbito del sida⁹.

INTERPRETACIÓN DE LOS RESULTADOS DE ENSAYOS CLÍNICOS BASADOS EN VRC

De la discusión anterior se deduce que hay indudables ventajas asociadas al uso de VRC. Sin embargo, si no se analiza el contexto en el que se han utilizado, la interpretación del efecto de una intervención puede ser errónea. Las VRC son, por tanto, un arma de doble filo cuyo manejo exige prudencia y un conocimiento de

TABLA 1. Variables de resultado en el estudio HOPE

	Intervención, n (%)	Control, n (%)	RR (IC del 95%)
IAM o ACV o muerte cardiovascular	651 (14)	826 (17,8)	0,78 (0,7-0,86)
IAM	459 (9,9)	570 (12,3)	0,8 (0,7-0,9)
ACV	156 (3,4)	226 (4,9)	0,68 (0,56-0,82)
Muerte cardiovascular	282 (6,1)	377 (8,1)	0,74 (0,64-0,87)

ACV: accidente cerebrovascular; IAM: infarto agudo de miocardio; IC: intervalo de confianza; RR: riesgo relativo.

TABLA 2. Variables de resultado en el estudio DREAM

	Intervención (n = 2.635), n (%)	Control (n = 2.634), n (%)	Tasa de riesgo (HR) (IC del 95%)
Diabetes de nueva aparición o muerte	306 (11,6)	686 (26)	0,4 (0,35-0,46)
Diabetes de nueva aparición	280 (10,6)	658 (25)	0,38 (0,33-0,44)
Muerte	30 (1,1)	33 (1,3)	0,9 (0,55-1,5)

HR: *hazard ratio*; IC: intervalo de confianza.

que las ambigüedades en algunos de los estudios que las utilizan no contribuyen a despejar. Desgraciadamente, en la literatura médica suele ser difícil concluir cuál es la motivación para el uso de VRC utilizadas en los ECA, sobre todo considerando que para los promotores de un fármaco puede ser más ventajoso centrarse en un resultado positivo basado en una VRC que entrar en debate sobre las precauciones a tener en cuenta para interpretar dicho efecto.

De esta manera, ha de ser el lector quien evalúe el riesgo de interpretación espuria del resultado de una intervención valorado mediante una VRC. El mayor riesgo lo constituyen las situaciones en que, habiendo un efecto claramente positivo en la VRC, se comprueba que dicho efecto aparece principalmente en un componente clínicamente poco importante, mientras que el efecto en componentes clínicamente relevantes es nulo o incluso negativo.

Montori et al⁵ han recomendado recientemente utilizar una guía de interpretación con objeto de valorar el riesgo de interpretación inexacta de resultados presentados mediante variables compuestas. Aunque en dicha guía no consta la motivación por la que se utilizó una VRC, lo cual podría representar una limitación de su uso en ciertos casos, constituye un primer paso útil para diferenciar los EC en que la interpretación del resultado basado en VRC es sencilla de aquellos en que es más compleja. La guía consta de tres preguntas básicas: ¿Darían los pacientes similar importancia a los componentes de la VRC? ¿Fueron similares las frecuencias con que ocurrieron los componentes? ¿Fueron similares los efectos de la intervención en cada uno de los componentes? Nuestra confianza en la valoración del efecto a partir de una VRC será progresivamente menor en la medida en que los componentes difieran en importancia, frecuencia y efecto del tratamiento.

Seguidamente proponemos dos ejemplos ilustrativos de VRC en las que el riesgo de interpretación espuria

de la intervención es mínimo y máximo respectivamente.

Ejemplo 1: estudio HOPE¹⁰

En ese estudio 9.297 pacientes con factores de riesgo cardiovascular fueron aleatorizados a recibir ramipril o placebo. Se usó la VRC «IAM o accidente cerebrovascular o muerte de origen cardiovascular». La tabla 1 muestra los resultados de la intervención en la VRC y en cada uno de sus componentes.

Ejemplo 2: estudio DREAM¹¹

Se aleatorizó a 5.269 pacientes sin enfermedad cardiovascular conocida y con intolerancia a la glucosa a rosiglitazona o placebo. Se utilizó la VRC «diabetes de nueva aparición o muerte». Los resultados se exponen en la tabla 2.

Importancia de los componentes (¿darían los pacientes similar importancia a los componentes de la VRC?)

Los componentes incluidos en una VRC deben ser de una importancia similar para los pacientes. En caso contrario, podría llegarse a conclusiones erróneas mezclando resultados muy diferentes. Si analizamos la heterogeneidad de las VRC en cuanto a la importancia de sus componentes en los ejemplos anteriores, la diferencia es fácil de apreciar. Mientras en el estudio HOPE hay un cierto gradiente de importancia entre los componentes (IAM o accidente cerebrovascular o muerte de origen cardiovascular), éste dista mucho del existente en el segundo ejemplo, en el que ambos componentes son muy diferentes en cuanto a su trascendencia para el paciente (diabetes de nueva aparición o muerte). Aunque esta estrategia es ciertamente subjetiva, puede servir como un primer paso para cla-

sificar los casos más problemáticos que, por otra parte, no son excepcionales en la literatura.

Frecuencia de eventos (¿Fue similar la frecuencia con la que ocurrieron los componentes?)

Cuanto mayor sea la diferencia en la frecuencia de eventos del grupo control entre componentes, más incertidumbre se tendrá sobre la aplicabilidad del efecto de la intervención en la VRC a dichos componentes. Mientras que en los componentes con una alta frecuencia de eventos la precisión del estimador del efecto será también alta, en los con escasa frecuencia de eventos la incertidumbre acerca de dicho estimador será mucho mayor, lo cual complica la interpretación del efecto. Esta estrategia nos sirve de guía para distinguir las situaciones que son más problemáticas que otras. Los ejemplos anteriores constituyen una situación ilustrativa. Mientras en el estudio HOPE la distribución de los eventos del grupo control varió entre el 4,9 y el 12,3%, en el estudio DREAM la heterogeneidad en la frecuencia de eventos fue sensiblemente mayor, concretamente el 1,3% para la variable muerte y un 25% para la variable «diabetes de nueva aparición».

Homogeneidad del efecto (¿Fue similar el efecto de la intervención en cada uno de los componentes?)

Es importante prestar atención al efecto de los diversos componentes para objetivar el grado de variabilidad entre ellos. El grado de variabilidad, en el caso de ser marcado, indicará que puede haber diferencias importantes entre los efectos sobre los componentes y cuestionar su evaluación conjunta. Al igual que en el caso anterior, el estimador del efecto de la intervención en los componentes (expresado en forma de riesgo relativo o de razón de tasas) es relativamente homogéneo en el estudio HOPE, con oscilaciones entre 0,7 y 0,8, y muy heterogéneo en el estudio DREAM, en el que varió de 0,38 a 0,9. Mientras en el primero se puede afirmar que el efecto de la intervención en la VRC es aplicable al resto de sus componentes, en el segundo la incertidumbre es máxima.

Reuniendo las tres preguntas anteriores podemos concluir que, mientras en el estudio HOPE es esperable que el efecto de la intervención en la VRC sea aplicable a sus componentes, en el estudio DREAM tenemos una incertidumbre muy importante sobre ello. Es más, la inferencia más prudente que podemos hacer en este último es que es plausible que la intervención tenga un efecto beneficioso en el riesgo de diabetes de nueva aparición. Por el contrario, no se puede establecer ninguna conclusión sobre el componente «mortalidad total».

A pesar de ello, los autores del estudio DREAM concluyen, «este gran ensayo clínico internacional, prospectivo y enmascarado, muestra que 8 mg de rosiglitazona al día, junto con recomendaciones higiénico-dietéticas, reduce sustancialmente el riesgo de diabetes o muerte en un 60% de los individuos con alto riesgo de diabetes». Si bien dicha aseveración es correcta, está claro que la afirmación de que dicha intervención reduce en un 60% el riesgo de diabetes o muerte induce al lector la impresión de un beneficio de la intervención en ambos componentes, una conclusión falsa que, por otra parte, tiende a magnificar el efecto del tratamiento observado en el ensayo.

USO DE VRC POTENCIALMENTE PROBLEMÁTICAS EN ENSAYOS CLÍNICOS CARDIOVASCULARES

Con objeto de explorar el uso de VRC potencialmente problemáticas en el ámbito cardiovascular en la vida real, se realizó un estudio de los ECA publicados en revistas de alto impacto que incluyen frecuentes estudios cardiovasculares⁶. El objetivo fue explorar la heterogeneidad de los componentes de las VRC primarias de los ECA elegibles en los tres dominios referidos en las normas prácticas de interpretación mencionadas: *a*) la importancia (relevancia clínica); *b*) la frecuencia de eventos, y *c*) la magnitud del efecto del tratamiento.

Para ello se llevó a cabo una revisión sistemática utilizando la base de datos MEDLINE de las revistas de medicina general y cardiología con mayor factor de impacto en el año 2003. En concreto se revisaron *Lancet*, *Annals of Internal Medicine*, *JAMA*, *New England Journal of Medicine*, *Circulation* y *European Heart Journal* desde el 1 de enero de 2002 hasta el 30 de junio de 2003. Se excluyeron las revistas pertenecientes al área cardiovascular pero del ámbito de experimentación básica (p. ej., *Circulation Research*). Se excluyeron además los estudios en que, aunque incluyeran alguna VRC, éstas estaban constituidas únicamente por componentes en relación con la seguridad o toxicidad de un fármaco o bien medidas paraclínicas o de laboratorio (*surrogate outcomes*). Asimismo, se excluyeron los estudios correspondientes a análisis de subgrupos que ignoraban la aleatorización.

Dos cardiólogos y 9 médicos internistas con formación en metodología de investigación clínica y epidemiología clasificaron de forma independiente las 72 variables de resultado encontradas que formaban parte de las VRC, en cinco categorías en orden decreciente de «importancia para el paciente»: I = muerte, II = crítica, III = mayor, IV = moderada y V = menor. El grupo de investigadores involucrado en la clasificación resolvió las discrepancias mediante discusión hasta llegar a un consenso de clasificación.

TABLA 3. Efecto aditivo de los componentes en la frecuencia de eventos del grupo control y en la magnitud del efecto de la intervención en la VRC según su importancia clínica

	Componentes más importantes considerados aisladamente	Añadiendo los componentes moderadamente importantes	Añadiendo los componentes menos importantes
Frecuencia de eventos del grupo control de la VRC, mediana (intervalo intercuartílico)	2,5% (0,8-6,1)	8,7% (5,2-14,5)	21,7% (13,1-33,2)
Efecto del tratamiento (RRR), mediana (intervalo intercuartílico)	13% (-14%-41%)	17% (-9%-36%)	24% (8%-42%)

RRR: reducción del riesgo relativo; VRC: variable de resultado combinada.

Se localizaron 242 ECA potencialmente elegibles, de los que 114 cumplieron criterios de inclusión y constituyeron la muestra para el análisis. En el 41% de los casos se reportó más de una VRC. La mayor parte de las veces, las VRC utilizadas estaban constituidas fundamentalmente por 2 (34%) o 3 componentes (39%); la mortalidad fue el componente más frecuente.

El estudio mostró que la inclusión de las variables de alta relevancia clínica (como mortalidad y otros eventos catalogados como «críticos») junto con variables de menor importancia relativa (categoría IV o V) en una VRC era común, apareciendo hasta en el 57% de los casos. En cuanto a la heterogeneidad del efecto de la intervención, se observó que en el 75% de los casos el efecto de la intervención en los componentes difería de forma moderada o importante. Lo mismo ocurrió con la frecuencia de eventos de los componentes. Globalmente, sólo un 14% de todas las VRC analizadas presentaban homogeneidad en los tres aspectos.

Se pudo demostrar, además, que en la mayoría de las ocasiones tanto el efecto de la intervención como la frecuencia de eventos estaban dominados por los componentes menos importantes, y el efecto en los componentes más importantes resultó clínicamente irrelevante. La tabla 3 muestra cómo aumentó de forma importante tanto la frecuencia de eventos del grupo control de la VRC como la magnitud del efecto del tratamiento conforme se añadían a la VRC componentes de menor relevancia clínica.

Por último, aunque la revisión sistemática se centró únicamente en las seis revistas de medicina general y cardiología de mayor factor de impacto en aquel momento, hay que mencionar que el uso de VRC es también común en nuestro ámbito, y se puede encontrar ejemplos de ello tanto en estudios aleatorizados como en observacionales¹².

En definitiva, el uso de VRC potencialmente problemáticas en cuanto a su interpretación en los ECA cardiovasculares es común. El mayor riesgo que comportan dichas VRC es favorecer la magnificación del beneficio real de la intervención al expresar el resultado de la intervención en términos, aparentemente plausibles, de mayor beneficio clínico. El lector de la literatura médica en general, y de la cardiológica en

particular, ha de ser especialmente cauto a la hora de interpretar los ECA cuyos resultados se expresan en forma de VRC.

CONCLUSIONES

Las VRC son un recurso metodológico frecuentemente utilizado. La finalidad más frecuentemente aducida para su uso es la de aumentar la eficiencia de los ensayos clínicos cuando se espera que el efecto estudiado en éstos sea pequeño. Pueden representar, asimismo, una medida del efecto neto global de una intervención y, ocasionalmente, ser útiles para evitar un sesgo por riesgos competitivos.

Es importante evaluar cuidadosamente los resultados de los estudios que utilizan VRC para evitar interpretaciones inadecuadas.

Cuando se utiliza una VRC, el significado clínico del efecto está relacionado con el grado de heterogeneidad de los componentes en tres dominios: importancia clínica relativa, magnitud del efecto y frecuencia de eventos. Cuanto mayor sea el grado de heterogeneidad en estos dominios, mayor será la incertidumbre sobre el significado clínico del efecto de la intervención.

En la literatura actual es frecuente el uso de VRC con un marcado gradiente de importancia clínica entre sus componentes y en las que la magnitud del efecto de la intervención predomina sobre los componentes de menor importancia. Estas circunstancias podrían favorecer la magnificación del beneficio real de las intervenciones que evalúan.

BIBLIOGRAFÍA

1. Moyé LA. Multiple analyses in clinical trials. Springer; 2003.
2. Neaton JD, Gray G, Zuckerman BD, Konstam MA. Key issues in end point selection for heart failure trials: composite end points. *J Card Fail.* 2005;11:567-75.
3. 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.
4. Ferreira-Gonzalez I, Permanyer-Miralda G, Busse JW, Bryant DM, Montori VM, Alonso-Coello P, et al. Methodologic discussions for using and interpreting composite endpoints are limited, but still identify major concerns. *J Clin Epidemiol.* 2007;60:651-7.

5. Montori VM, Permyner-Miralda G, Ferreira-Gonzalez I, Busse JW, Pacheco-Huergo V, Bryant D, et al. Validity of composite end points in clinical trials. *BMJ*. 2005;330:594-6.
6. Ferreira-Gonzalez I, Busse JW, Heels-Ansdell D, Montori VM, Akl EA, Bryant DM, et al. Problems with use of composite end points in cardiovascular trials: systematic review of randomised controlled trials. *BMJ*. 2007;334:786.
7. Wallentin L, Goldstein P, Armstrong PW, Granger CB, Adgey AA, Arntz HR, et al. Efficacy and safety of tenecteplase in combination with the low-molecular-weight heparin enoxaparin or unfractionated heparin in the prehospital setting: the Assessment of the Safety and Efficacy of a New Thrombolytic Regimen (ASSENT)-3 PLUS randomized trial in acute myocardial infarction. *Circulation*. 2003;108:135-42.
8. Keech A, Simes RJ, Barter P, Best J, Scott R, Taskinen MR, et al. Effects of long-term fenofibrate therapy on cardiovascular events in 9795 people with type 2 diabetes mellitus (the FIELD study): randomised controlled trial. *Lancet*. 2005;366:1849-61.
9. Del Amo J, Perez-Hoyos S, Moreno A, Quintana M, Ruiz I, Cisneros JM, et al. Trends in AIDS and mortality in HIV-infected subjects with hemophilia from 1985 to 2003: the competing risks for death between AIDS and liver disease. *J Acquir Immune Defic Syndr*. 2006;41:624-31.
10. Yusuf S, Sleight P, Pogue J, Bosch J, Davies R, Dagenais G. Effects of an angiotensin-converting-enzyme inhibitor, ramipril, on cardiovascular events in high-risk patients. The Heart Outcomes Prevention Evaluation Study Investigators. *N Engl J Med*. 2000;342:145-53.
11. DREAM (Diabetes REduction Assessment with ramipril and rosiglitazone Medication) Trial Investigators, Gerstein HC, Yusuf S, Bosch J, Pogue J, Sheridan P, Dinccag N, et al. Effect of rosiglitazone on the frequency of diabetes in patients with impaired glucose tolerance or impaired fasting glucose: a randomised controlled trial. *Lancet*. 2006; published online Sept 15. DOI:10.1016/S0140-6736(06)69420-8.
12. Aldamiz-Echevarría B, Muñoz J, Rodríguez-Fernández JA, Vidán-Martínez L, Silva-César M, Lamelo-Alfonsín F, et al. Ensayo clínico aleatorizado y controlado para valorar una intervención por una unidad de hospitalización domiciliaria en la reducción de reingresos y muerte en pacientes dados de alta del hospital tras un ingreso por insuficiencia cardíaca. *Rev Esp Cardiol*. 2007;60:914-22.