

ANÁLISIS DE SUBGRUPOS



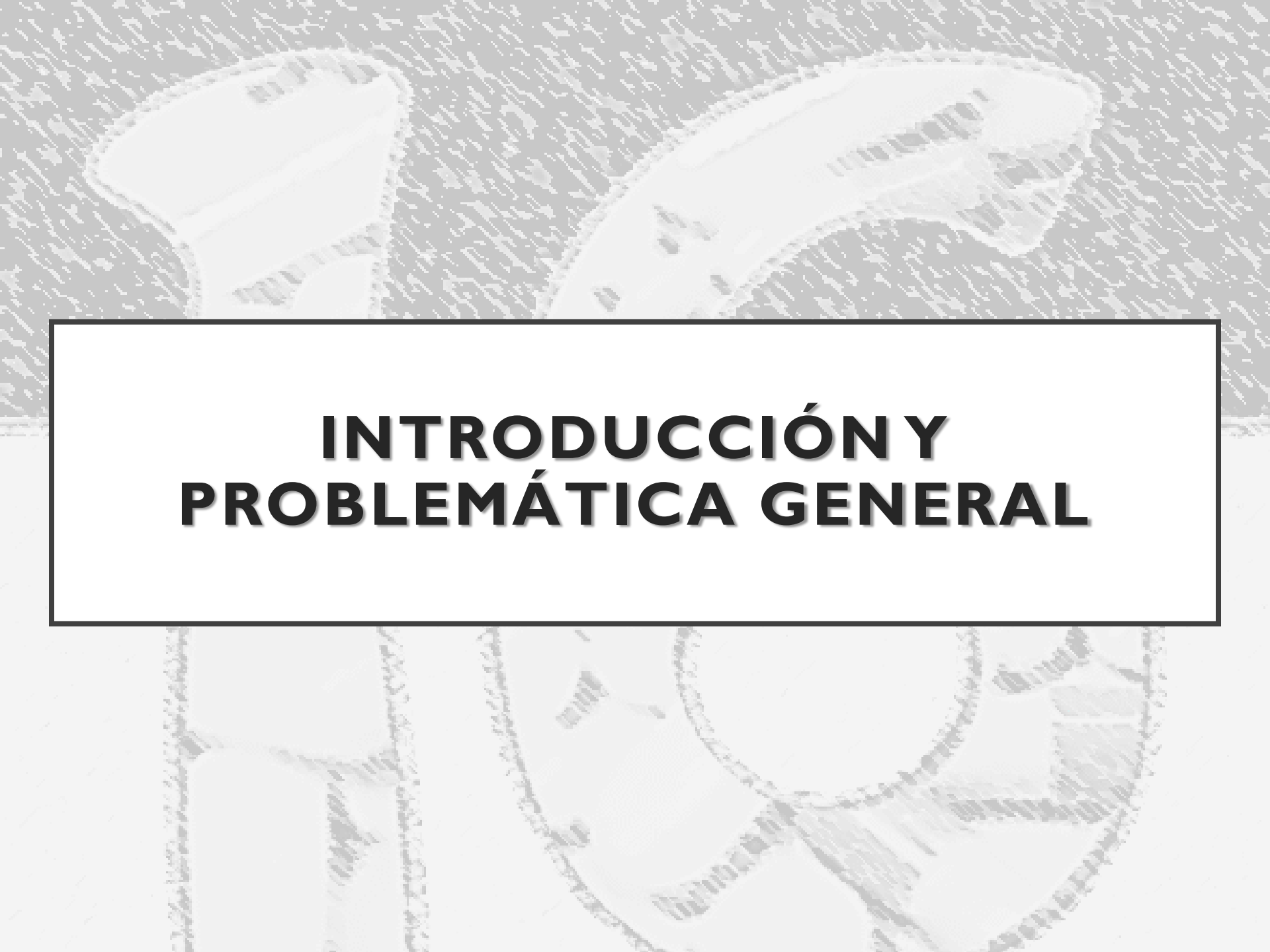
Silvia Fénix Caballero

H. U. Puerto Real

Servicio Andaluz de Salud



**Curso de Evaluación y
Selección de Medicamentos**

The background of the slide features a faint, stylized illustration of two hands, one from the top and one from the bottom, gently cradling a globe. The hands are rendered in a sketchy, charcoal-like style. The globe shows some geographical details like continents and latitude lines. The entire background has a fine, cross-hatched texture.

INTRODUCCIÓN Y PROBLEMÁTICA GENERAL

CASOS QUE SE PRESENTAN EN LA PRÁCTICA DE LAS EVALUACIONES

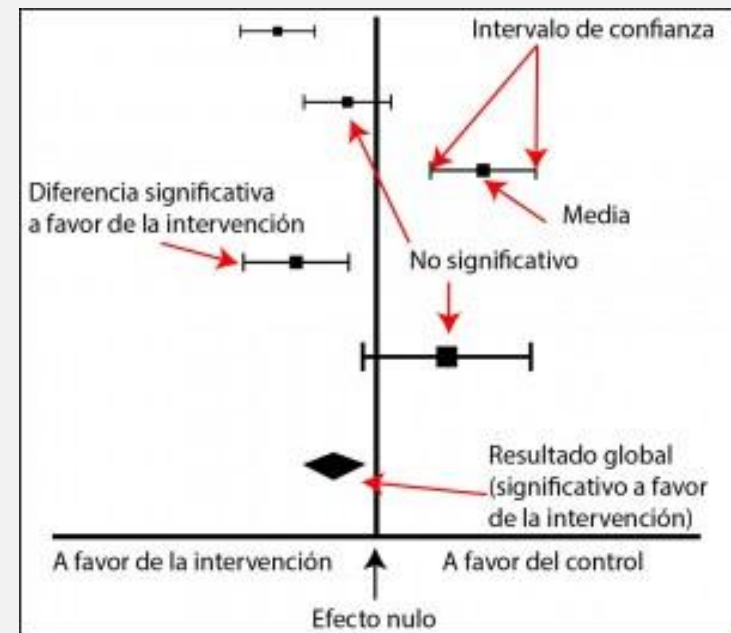
1. ECA globalmente **POSITIVO**, pero...

- Mezcla poblaciones muy diversas
- Existen subgrupos con eficacia dudosa

Fármacos muy ineficientes: ¿seleccionar pacientes y maximizar eficiencia?

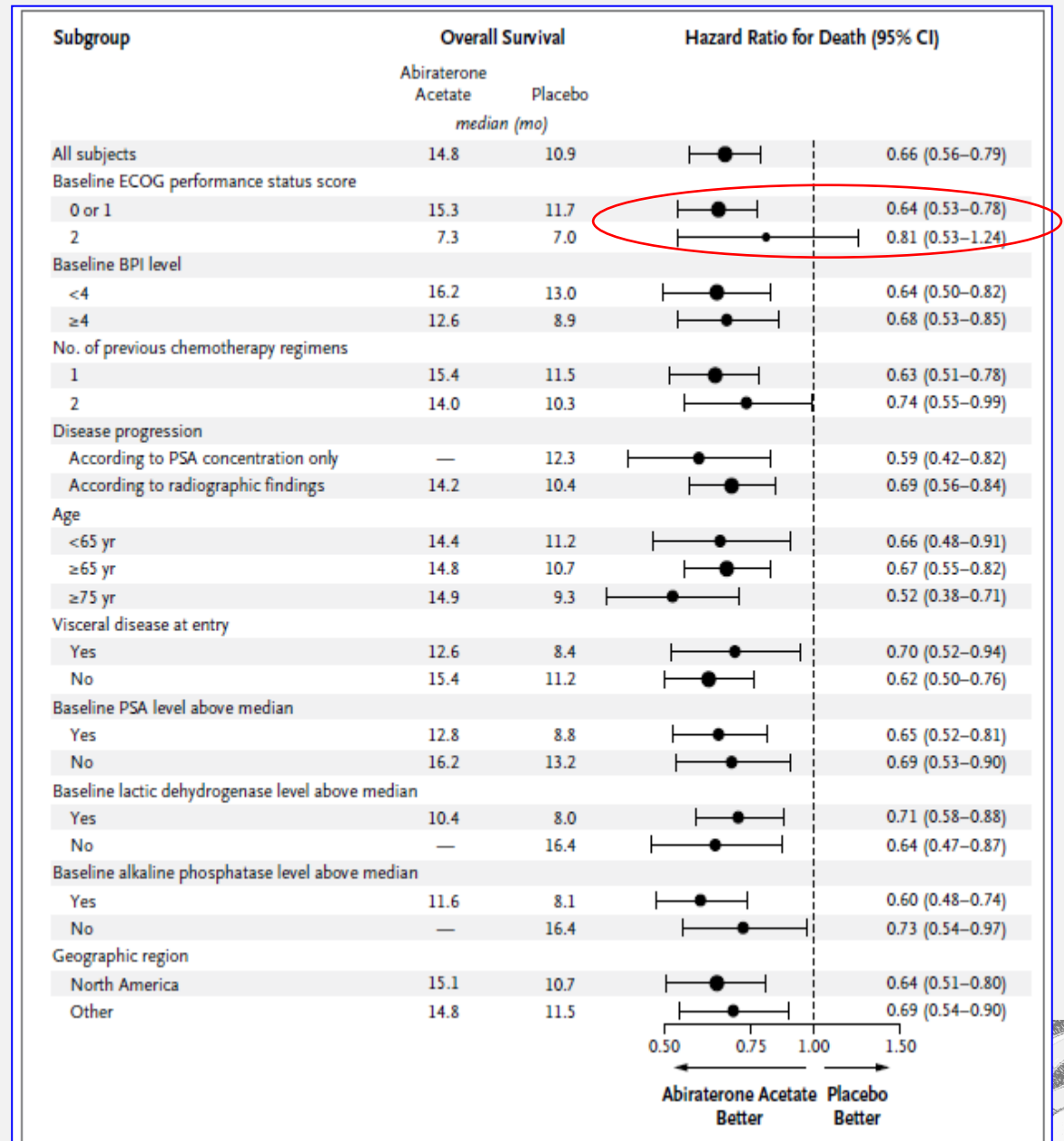
2. ECA globalmente **NEGATIVO**, pero...

- Subgrupo con diferencias estadísticamente significativas.



FOREST-PLOT

N ENGL J MED 364;21 NEJM.ORG MAY 26, 2011



N Engl J Med. 1978 Jul 13;299(2):53-9.

A randomized trial of aspirin and sulfinpyrazone in threatened stroke. The Canadian Cooperative Study Group.

[No authors listed]

Abstract

Five hundred and eighty-five patients with threatened stroke were followed in a randomized clinical trial for an average of 26 months to determine whether aspirin or sulfinpyrazone, singly or in combination, influence the subsequent occurrence of continuing transient ischemic attacks, stroke or death. Eighty-five subjects went on to stroke, and 42 died. Aspirin reduced the risk of continuing ischemic attacks, stroke or death by 19 per cent (P less than 0.05) and also reduced risk for the "harder," more important events of stroke or death by 31 percent (P less than 0.05), but this effect was sex-dependent: among men, the risk reduction for stroke or death was 48 per cent (P less than 0.005), whereas no significant trend was observed among women. For sulfinpyrazone, no risk reduction of ischemic attacks was observed, and the 10 per cent risk reduction of stroke or death was not statistically significant. No overall synergism or antagonism was observed between the two drugs. We conclude that aspirin is an efficacious drug for men with threatened stroke.

Durante años, muchos médicos privaron el uso de aspirina en mujeres con enfermedad cerebrovascular.

ANÁLISIS DE SUBGRUPOS QUE RESULTARON FALSOS

- La aspirina es inefectiva como prevención secundaria de AVE en mujeres^{2,26,27}
- La terapia antihipertensiva es inefectiva como prevención cardiovascular primaria en mujeres^{30,31}
- La terapia antihipertensiva es inefectiva o dañina en ancianos³⁰
- IECA no reducen la mortalidad y hospitalización en pacientes con ICC que toman AAS³²
- Los beta bloqueadores son inefectivos tanto en IAM de pared inferior como en ancianos^{29,33}
- La trombolisis es inefectiva en IAM de > 6 h³⁴
- El tamoxifeno es inefectivo en mujeres con cáncer de mama < 50 años³⁵
- El amlodipino reduce la mortalidad en pacientes con ICC de origen no isquémico pero no en los de origen isquémico³⁶



FALSOS -

- Habiendo **Diferencias Reales** entre los subgrupos los ECAs o las RS **NO** son capaces de identificarlos
- (41-66%)

Ocurre cuando :

- ✧ **N pequeña**
- ✧ **Diferencia del Efecto ↓**

FALSOS +

- **Muestra un Efecto** significativamente distinto cuando en realidad **NO** existe
- (7-21%)

Ocurre cuando:

- ↑ **n° de subgrupos** ↑ la probabilidad de encontrar diferencias que no existen sólo por azar



Limitaciones del análisis de subgrupos:

- ✓ Determinaciones de resultados suplementarias:
 - ↑ Posibilidad de obtener diferencias debidas al azar

↑ α

- ✓ División de la muestra (al límite del poder estadístico)
 - No detección de diferencias que sí existen

↑ β

EL PROBLEMA ESENCIAL DEL ANÁLISIS DE SUBGRUPOS

Por multiplicidad de mediciones

↑ Error tipo I (alfa, falso +)
Aparecen diferencias que no existen

α

Por reducción de tamaño muestral

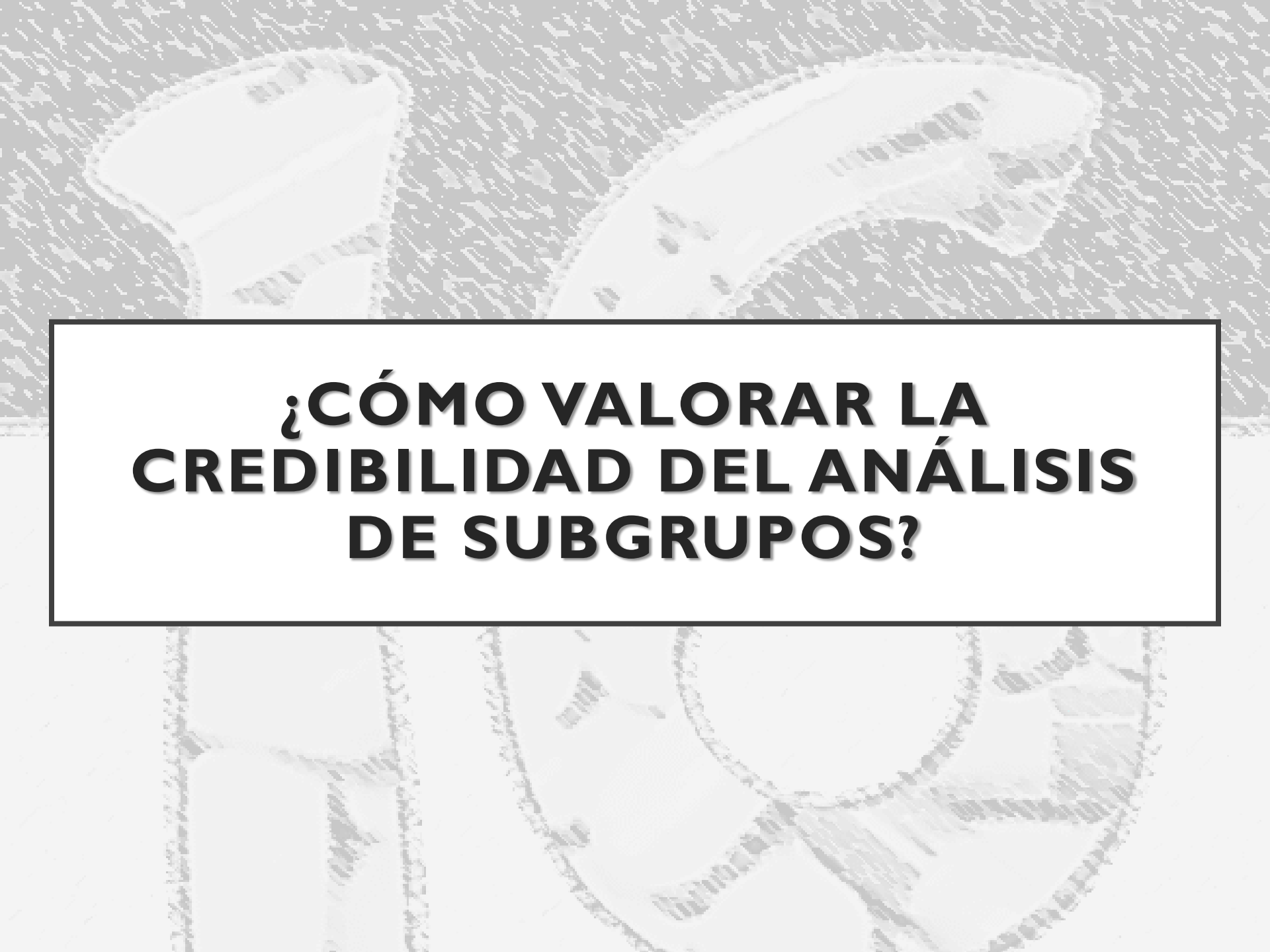
↑ Error tipo II (beta, falso -)
No se encuentran diferencias que sí existen

β

↓ error:
Descartando falsos positivos:
plausibilidad biológica, consistencia



↓ error:
Menos exigentes con la interacción:
 $p < 0,1$
en lugar de $< 0,05$

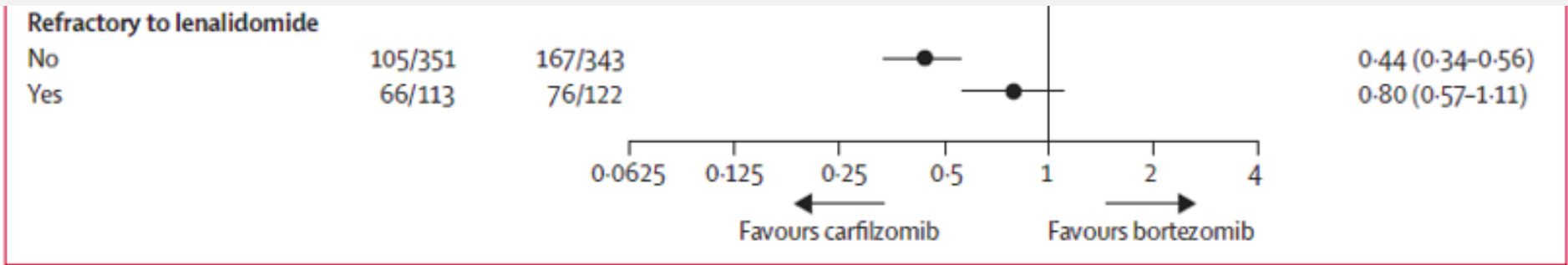


¿CÓMO VALORAR LA CREDIBILIDAD DEL ANÁLISIS DE SUBGRUPOS?

PREGUNTA

Dimopoulos MA, et al. *Carfilzomib and dexamethasone vs. bortezomib and dexamethasone for patients with relapsed or refractory multiple myeloma (ENDEAVOR)*. *Lancet Oncol.* 2016;17(1):27-38.

- La refractariedad al tratamiento previo con lenalidomida, ¿es un criterio útil para predecir menor beneficio en SLP de carfilzomib?



- ... Esta aparente diferencia de eficacia, ¿se debe al azar o existe realmente?



How to Use a Subgroup Analysis

Users' Guides to the Medical Literature

Xin Sun, PhD; John P. A. Ioannidis, MD, DSc; Thomas Agoritsas, MD; Ana C. Alba, MD; Gordon Guyatt, MD, MSc

Analizar con criterio:

- **INTERACCIÓN** ¿Puede **el azar** explicar las diferencias entre los subgrupos?
+
- **PREESPECIFICACIÓN** ¿Fue la diferencia de subgrupo procedente **de un pequeño número** de análisis justificados? ¿Se especificó **previamente**?
- **PLAUSIBILIDAD BIOLÓGICA** ¿Hay una **hipótesis** que apoye el efecto de los subgrupos? ¿Estaba descrita con anterioridad?
- **CONSISTENCIA** ¿Es la diferencia en los subgrupos **consistente** en diferentes estudios?

INTERACCIÓN

- Diferencia estadísticamente significativa entre diferentes subgrupos → **INTERACCIÓN**

Probabilidad (p) de que la diferencia entre subgrupos se deba al azar

Análisis estadístico de interacción

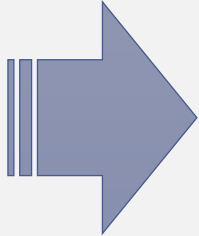


Pruebas estadísticas: determinan hasta qué punto las diferencias pueden deberse al azar

SIGNIFICACIÓN ESTADÍSTICA

Límite de la significación $\rightarrow p < 0,05$

Generalmente



Subgrupos: $\downarrow$ muestra $\rightarrow$ este límite podría no detectar diferencias probablemente relevantes

0,1

Si no existe significación estadística en el resultado del subgrupo → ausencia de beneficio comparativo



Al dividir la muestra, la potencia estadística disminuye

Podría ser insuficiente para detectar diferencias

DIFERENCIAS INTERGRUPALES

No se trata de
comprobar si existe
eficacia en todos los
subgrupos



**existen diferencias
de eficacia entre
subgrupos**

DIFERENCIAS INTERGRUPALES

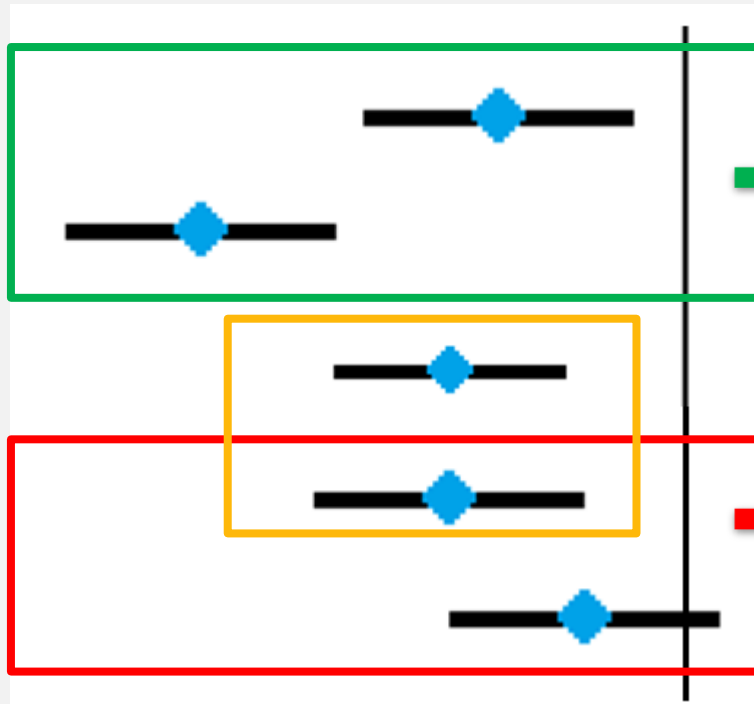


Diferencia de beneficio → AZAR ??

Si las diferencias (considerando IC95%) son compatibles con el azar →
factor estudiado no influye en el resultado

**Sólo con diferencias intergrupales → podemos
atender a los resultados específicos del subgrupo
estudiado**

Identificar subgrupos a analizar o desechar



No existe solapamiento:
 $p < 0,05$

Extremo alcanza valor central:
 $p > 0,2$

NO INTERACCIÓN

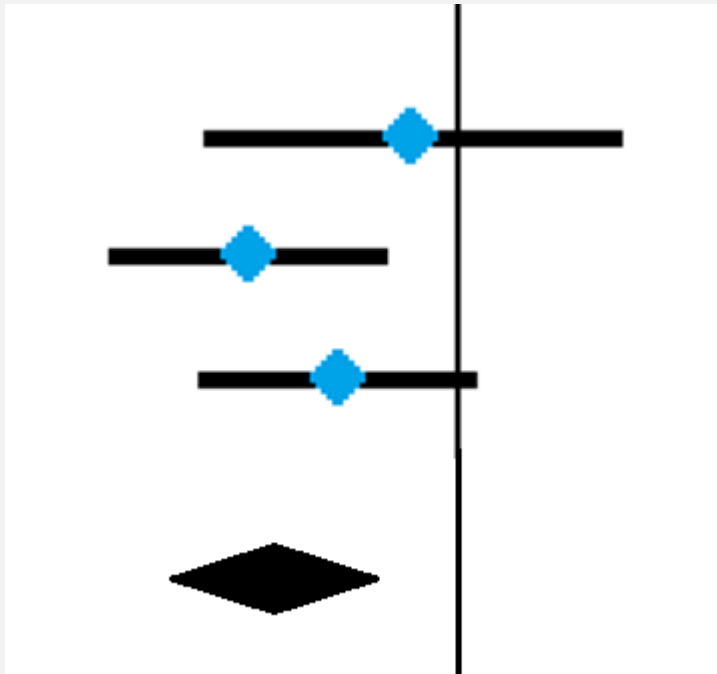
Existe diferencia, pero no es significativa

- No puede rechazarse H_0
- Decisión terapéutica basada en resultado global del estudio (no en estos grupos)

RESULTADO GLOBAL DEL ESTUDIO POSITIVO

ESCENARIOS

I. NO INTERACCIONES SIGNIFICATIVAS ENTRE SUBGRUPOS

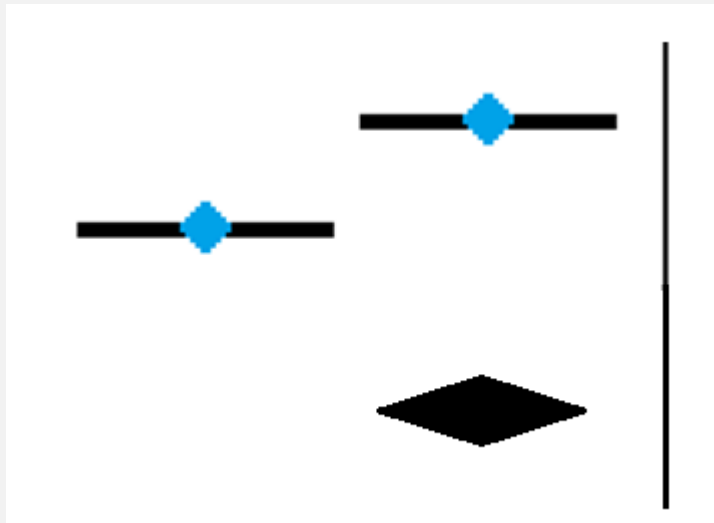


Resultados subgrupos consistente con el **global del estudio**, haya o no significación estadística intragrupal

RESULTADO GLOBAL DEL ESTUDIO POSITIVO

ESCENARIOS

2. INTERACCIÓN INTERGRUPAL E INTRAGRUPAL



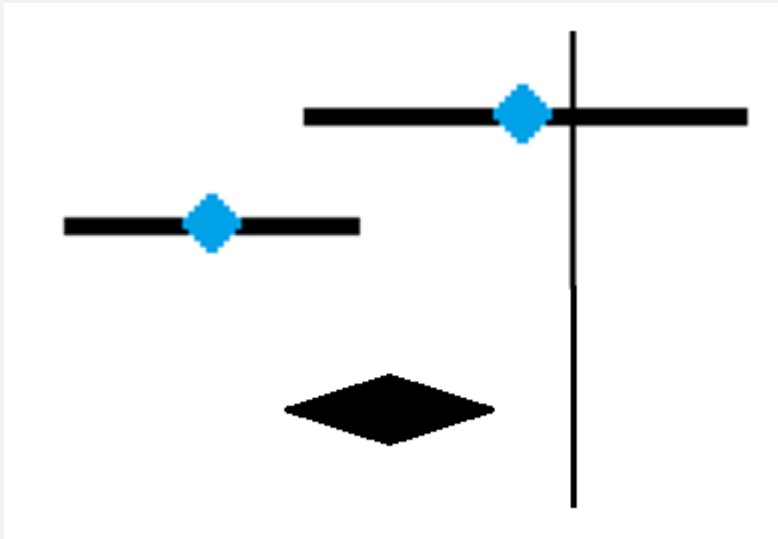
Consistencia con el resultado global, aunque el **beneficio es más probable** en un subgrupo que en otro

DIFERENCIA CUANTITATIVA

RESULTADO GLOBAL DEL ESTUDIO POSITIVO

ESCENARIOS

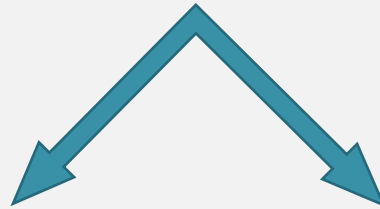
3. INTERACCIÓN INTERGRUPAL



Existe interacción, y uno de los subgrupos no presenta resultado estadísticamente significativo

**VALORAR CRITERIOS
ADICIONALES**

CALCULADORAS



Calculadora de interacción de
Joaquín Primo

2 subgrupos

P de homogeneidad de meta-
análisis

>2 subgrupos

2 subgrupos

Calculadora de interacción (Joaquín Primo)

RR-OR-HR

	SUBGRUPO 1	SUBGRUPO 2		
RR, OR o HR	0,67	0,88	Diferencia ln (Sub1-Sub2)	-0,2726
Límite inferior IC 95%	0,46	0,71	Error estándar	0,2206
Límite superior IC 95%	0,98	1,08	Límite inferior IC 95%	-0,7051
			Límite superior IC 95%	0,1598
ln(RR, OR o HR)	-0,4005	-0,1278	Z	-1,2358
Límite inferior IC 95%	-0,7765	-0,3425	P interacción	0,2165
Límite superior IC 95%	-0,0202	0,0770		0,2165
Amplitud IC 95%	0,7563	0,4195		
Error estándar	0,1929	0,1070	Razón de RR, OR o HR	0,76
			Límite inferior IC 95%	0,49
			Límite superior IC 95%	1,17

Altman DG, Bland JM. BMJ 2003;326:219.

J. Primo. Hospital de Sagunto

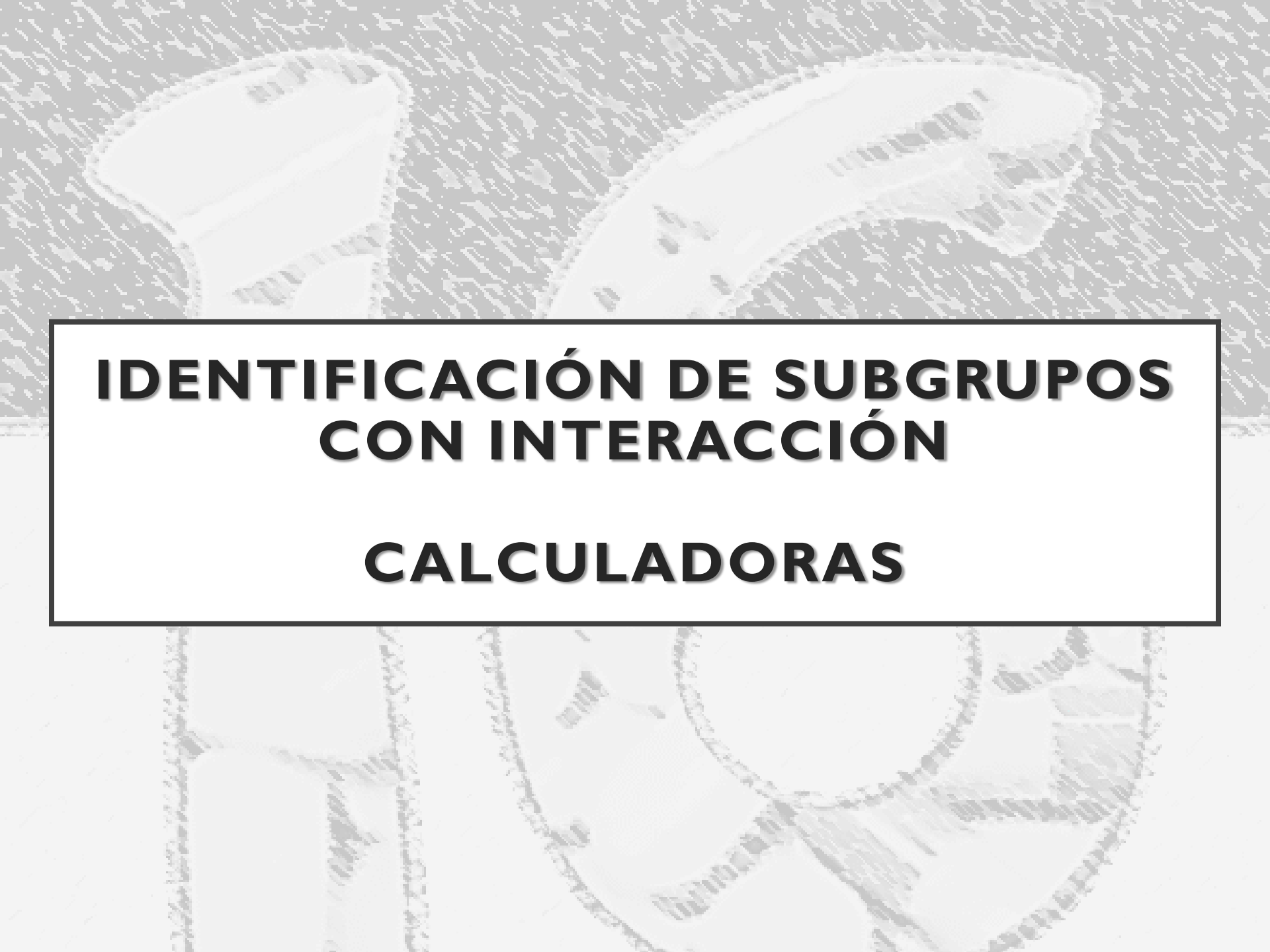
MEDIAS

PROPORCIONES

+ 2 subgrupos

p de interacción meta-análisis
(Calculadora Joaquín Primo)

Nº de estudios incluidos: 3								
Referencia (autor, año)	Nº	Hazard Ratio	V (ln HR)	IC 95% Hazard Ratio	z	p	Peso (%) EF	Peso (%) EA
	1	0,49	0,1344	0,24 a 1,01	1,95	0,05168	33,70	33,70
	2	0,55	0,1434	0,26 a 1,16	1,58	0,11440	31,59	31,59
	3	0,52	0,1305	0,26 a 1,06	1,81	0,07027	34,71	34,71
	4							
	5							
	6							
		Hazard Ratio combinado		IC 95%	z	p		
Efectos fijos (inverso de la varianza):		0,52		0,34 a 0,79	3,08	0,00205		
Efectos aleatorios (DerSimonian-Laird):		0,52		0,34 a 0,79	3,08	0,00205		
			p			IC 95%		
Q de heterogeneidad:		0,05	0,97624	I ² :	0%	0% a 0%,	Rosenthal:	8

The background of the slide features a grayscale, textured image of two hands cupping a globe. The hands are positioned at the top and bottom, with the globe in the center. The texture is grainy and sketch-like.

IDENTIFICACIÓN DE SUBGRUPOS CON INTERACCIÓN

CALCULADORAS

ELOTUZUMAB

Brazos de tratamiento

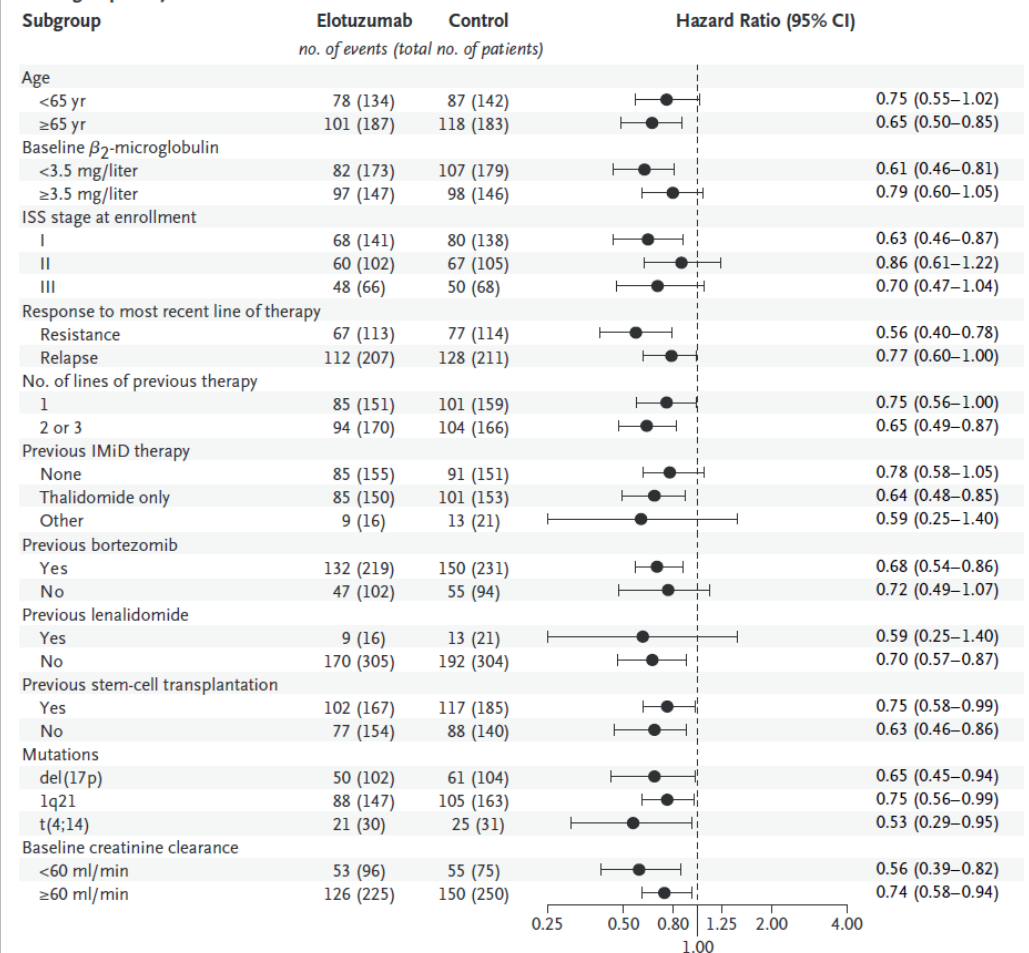
Experimental: Elotuzumab-Lenalidomida-dexametasona

Control: Lenalidomida-dexametasona

Indicación

Mieloma múltiple

B Subgroup Analyses

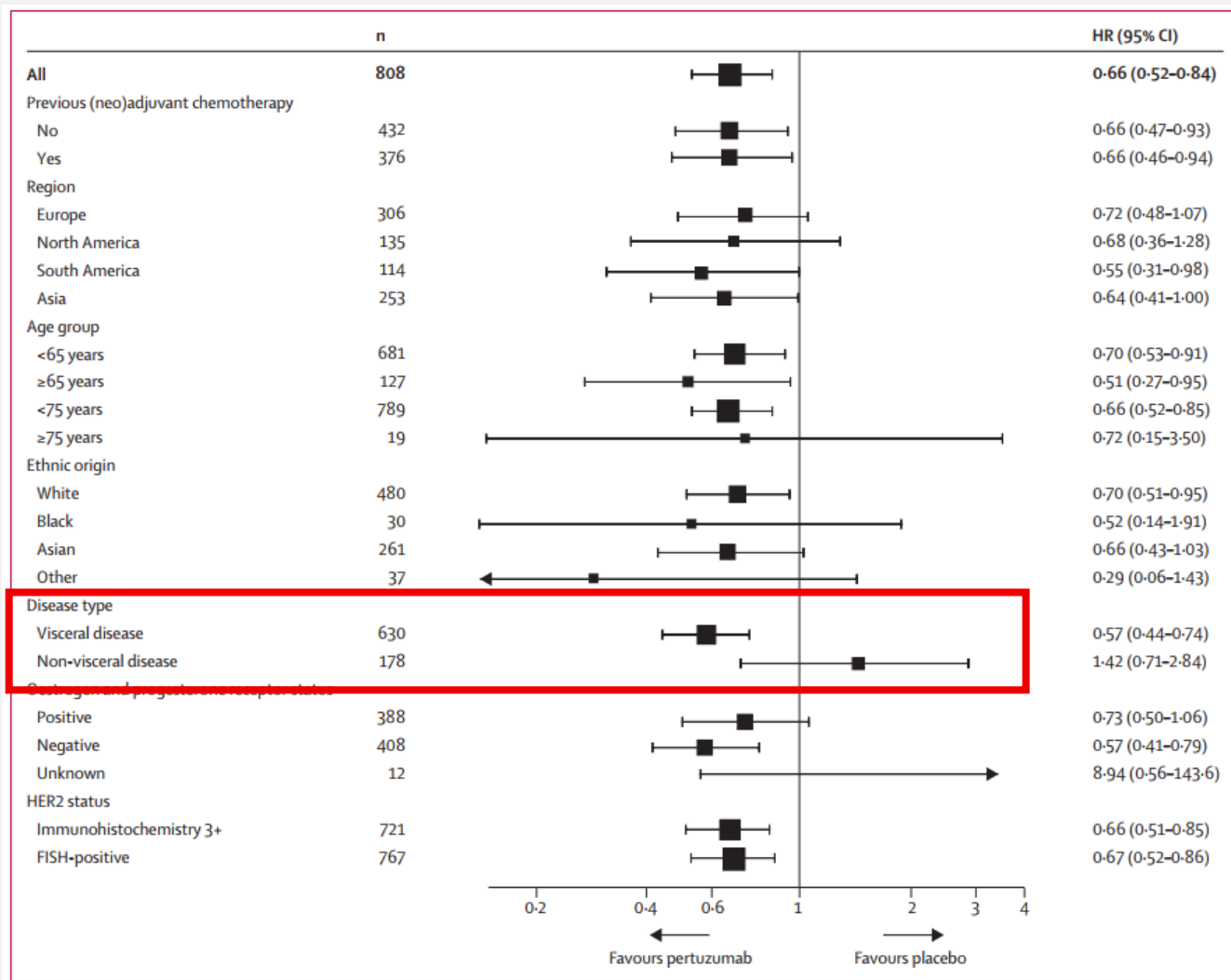


Los subgrupos se solapan

NO DIFERENCIAS

PERTUZUMAB

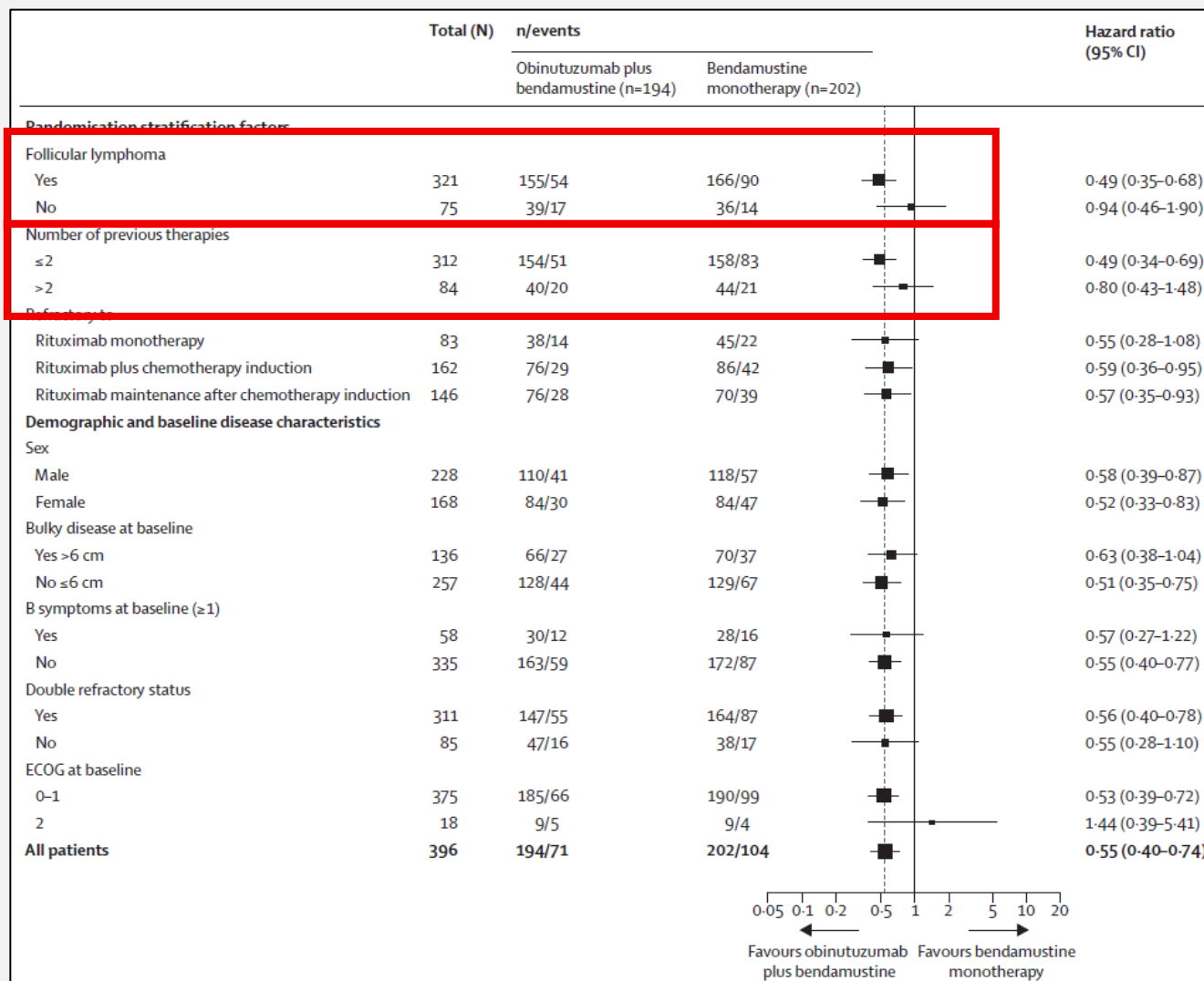
Brazos de tratamiento	<u>Experimental</u> : Pertuzumab, trastuzumab y docetaxel <u>Control</u> : Placebo, trastuzumab y docetaxel
Indicación	Cáncer de mama



P=0,0157

OBINUTUZUMAB

Brazos de tratamiento	Experimental: Obinutuzumab + bendamustina Control: bendamustina
Indicación	Linfoma no Hodgkin

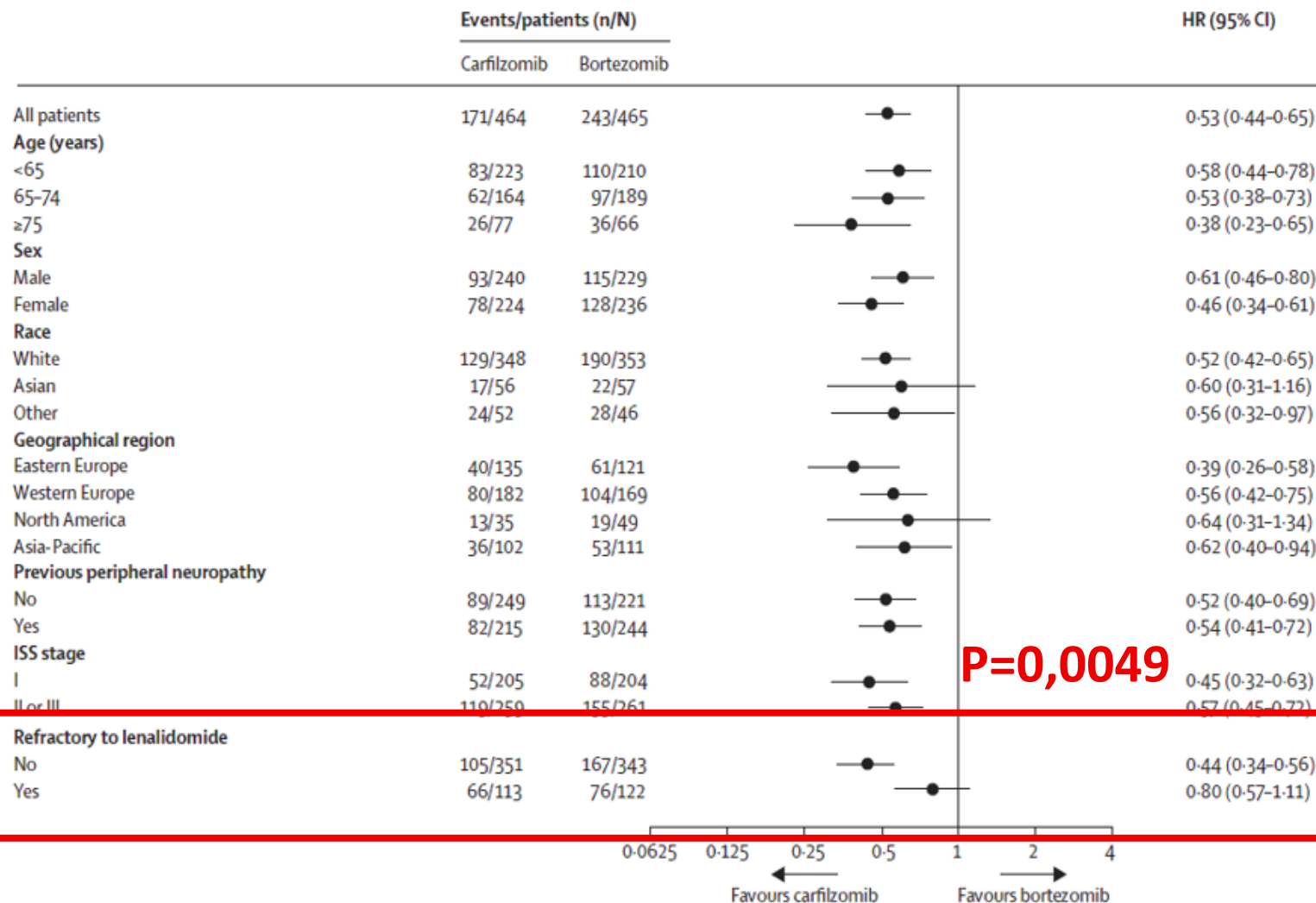


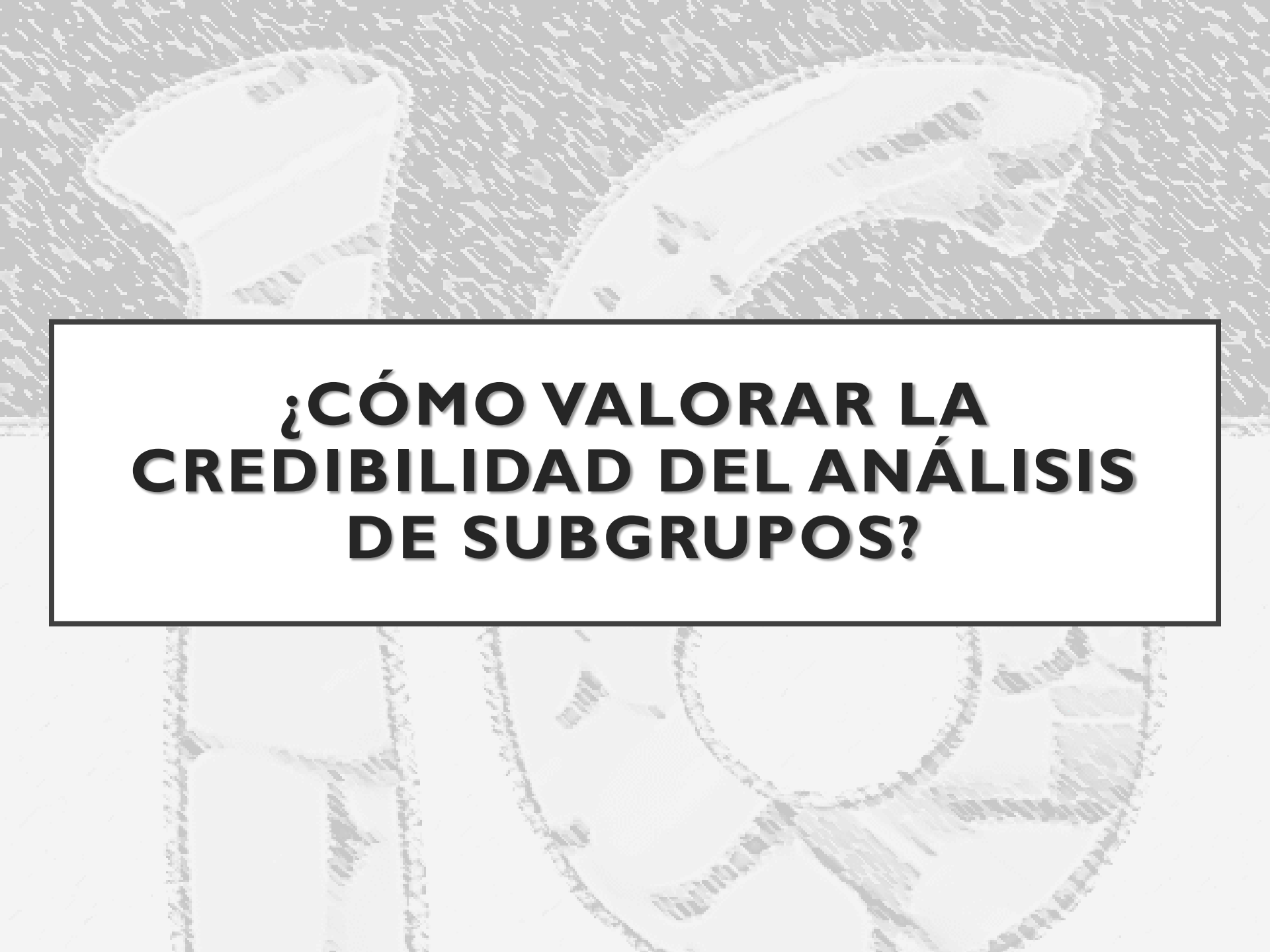
P=0,1030

P=0,1773

CARFILZOMIB

Brazos de tratamiento	Experimental: Carfilzomib-dexametasona Control: Bortezomib-dexametasona
Indicación	Mieloma múltiple





¿CÓMO VALORAR LA CREDIBILIDAD DEL ANÁLISIS DE SUBGRUPOS?

VOLVAMOS A LA CUESTIÓN

Refractory to lenalidomide

No

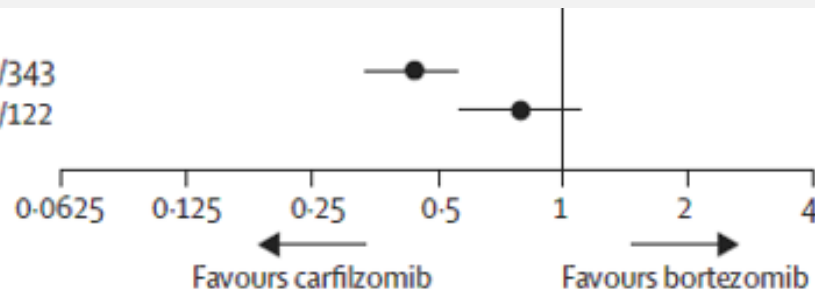
105/351

167/343

Yes

66/113

76/122



$p=0,0049$

A parte de la **EXISTENCIA DE DIFERENCIA** intergrupar (interacción)

Qué más podríamos valorar para tener mayor certeza y aplicar esta posible diferencia al posicionamiento del fármaco?

How to Use a Subgroup Analysis

Users' Guides to the Medical Literature

Xin Sun, PhD; John P. A. Ioannidis, MD, DSc; Thomas Agoritsas, MD; Ana C. Alba, MD; Gordon Guyatt, MD, MSc

Analizar con criterio:

- **INTERACCIÓN** ¿Puede **el azar** explicar las diferencias entre los subgrupos?
+
- **PREESPECIFICACIÓN** ¿Fue la diferencia de subgrupo procedente **de un pequeño número** de análisis justificados? ¿Se especificó **previamente**?
- **PLAUSIBILIDAD BIOLÓGICA** ¿Hay una **hipótesis** que apoye el efecto de los subgrupos? ¿Estaba descrita con anterioridad?
- **CONSISTENCIA** ¿Es la diferencia en los subgrupos **consistente** en diferentes estudios?

PREESPECIFICACIÓN

1. Subgrupo pre-especificado o factor de estratificación
2. Pre-especificado, pero los grupos se han dividido de forma distinta a la preespecificada (*menor evidencia*)

El **análisis “post hoc”** aumenta la probabilidad de encontrar falsas diferencias



Nº
subgrupos
limitados



A ↑ nº de subgrupo: ↑ riesgo de detectar diferencias que no existen, sólo por azar.

PLAUSIBILIDAD BIOLÓGICA

Existencia de una **hipótesis plausible** para explicar las diferencias encontradas estadísticamente.

Mayor credibilidad si el efecto de subgrupo tiene un **fundamento que lo apoya**:

- ✗ la racionalidad biológica
- ✗ fisiopatológica
- ✗ experimentación en animales

Los análisis de subgrupos con efectos que **NO** explicables con el **conocimiento biológico existente**, con frecuencia son **irreales**

(Yusuf S, Wittes J, Probstfield J, 1991).

Ej: signos del Zodíaco en el ensayo ISIS-2

Aspirina vs. placebo en IAM

Oh, SORPRESA !!

¡Oh, no! Soy Géminis: Interpretación de los análisis de subgrupos

Esa amarga lamentación
debieron exclamar los nacidos
bajo el signo zodiacal de géminis
que leyeron el [análisis de
subgrupos del estudio ISIS-2](#) que

demostraba ausencia de efecto
de la aspirina frente a placebo
tras un infarto agudo de
miocardio, en los pacientes
pertenecientes a los signos
zodiacales de Libra y Géminis y
una diferencia estadísticamente
significativa con respecto a los
que pertenecían a cualquiera de
los otros signos. ¿Tendría
sentido que los medicamentos fueran más o menos eficaces en los pacientes en
función de su signo del zodiaco? Apparently la medicina personalizada no llega



Ej: signos del Zodíaco en el ensayo ISIS-2

Table 1. Subgroup Analysis of the Second International Study of Infarct Survival

	No./Total (%) of Patients		Relative Risk (95% CI)	Relative Risk Reduction or Increase, %
	Aspirin	Placebo		
Vascular mortality in all patients	804/8587 (9.4)	1016/8600 (11.8)	0.79 (0.73-0.87)	20.7 Reduction
Gemini or Libra	150/1357 (11.1)	147/1442 (10.2)	1.08 (0.87-1.34)	8.4 Increase
Other astrological signs	654/7228 (9.0)	868/7187 (12.1)	0.75 (0.68-0.82)	25.4 Reduction

P de interacción 0,003

Schulz KT, Grimes DA. *Lancet* 2005; 365: 1657-61



En ISIS-2 aspirina
era ineficaz en
pacientes de signo
del zodiaco **Libra** y

**¿Qué habríamos
pensado si en lugar
del zodiaco hubiesen
empleado una
variable fisiológica?**

CONSISTENCIA

Comportamiento en

- **Otras variables de resultados relacionadas**
- **Otros fármacos con mismo mecanismo acción**
- **Otros estudios del mismo fármaco en otra indicación**

Si un análisis de subgrupos muestra un efecto consistente en **otras variables de resultado** con fisiopatología similar, el efecto es **más creíble**

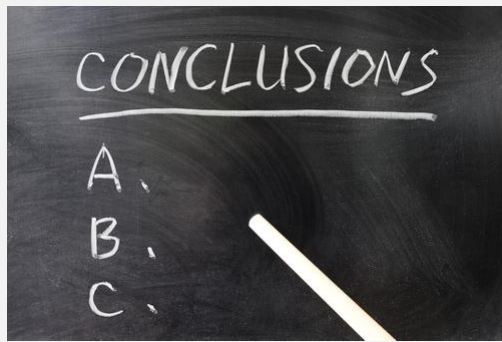
¿VALORACIÓN FINAL?

Analizar con criterio:

-Estadístico: Test de **INTERACCIÓN**

+

-Requisitos adicionales **CLAVES**



MAYOR CREDIBILIDAD SI.....

- ✧ Se especifica la hipótesis **a priori**
- ✧ Nº **reducido** subgrupos
- ✧ La interacción es **estadísticamente significativa**
- ✧ **Consistente** entre los estudios
- ✧ Existe **plausibilidad biológica**

Agradecimientos:
Emilio Alegre del Rey
Laila Abdel-Kader Martín



TALLER CHECKLIST

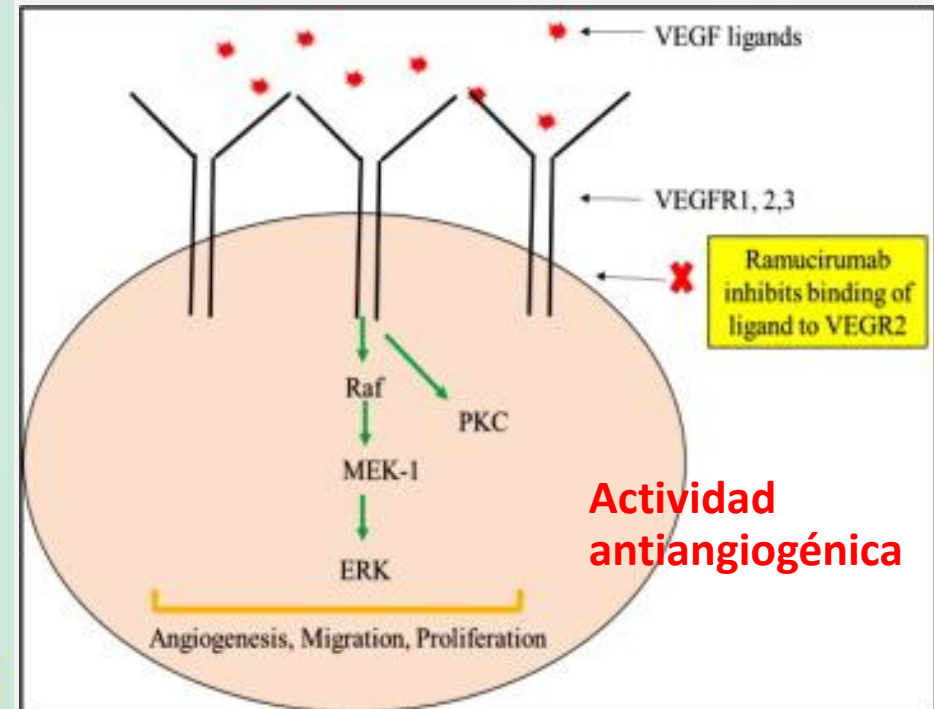
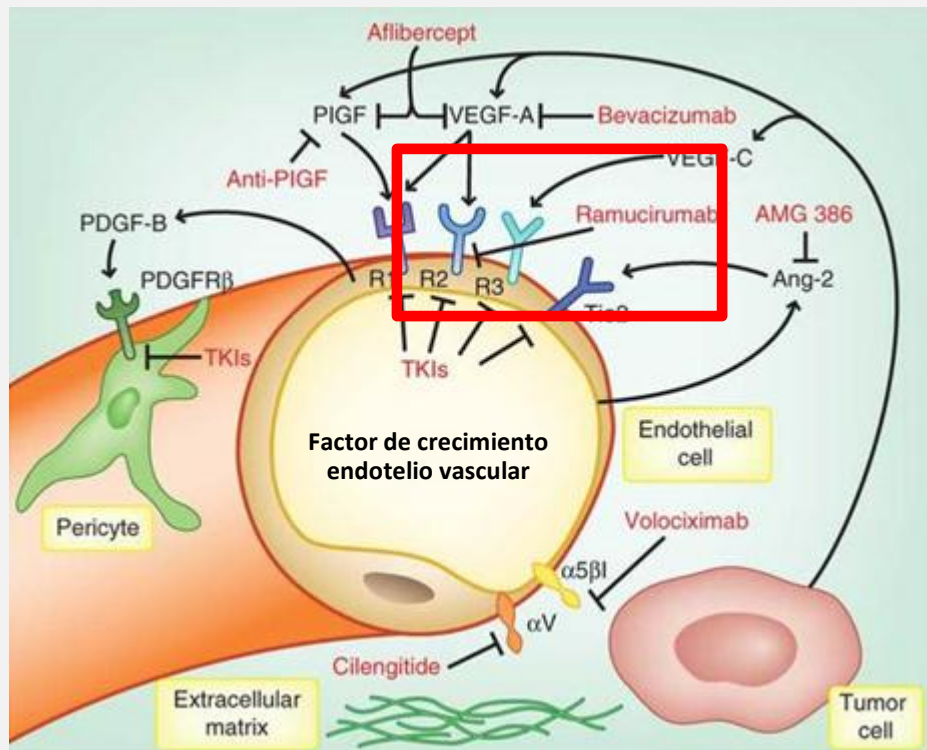


RAMUCIRUMAB ($\pm$ paclitaxel)

INDICACIÓN

Ca gástrico avanzado o adenocarcinoma de unión gastroesofágica tras QT basada en Pt y fluoropirimidina.

MECANISMO ACCIÓN





RAMUCIRUMAB ($\pm$ paclitaxel)

ESCENARIO

Opciones de tratamiento (2ª línea):

Irinotecán, paclitaxel y docetaxel

→ Obtienen mejoras en SG en comparación con el mejor tratamiento de soporte.

Coste ramucirumab

			VARIABLE	Eficacia de A	Eficacia de B	Diferencia (IC95%)	Coste incremental	CEI por paciente y año de vida ganado (AVG)
REGARD	RAMU (A) vs. PBO (B)	SG		5.2 meses	3.8 meses	1.4 meses (0.117 años)	11.920,888 €	101.887,93 €
RAINBOW	RAMU+PACL (A) vs. PACL (B)	SG		9.6 meses	7.4 meses	2.2 meses (0.183 años)	26.884,41 €	146.909,34 €

RAMUCIRUMAB ($\pm$ paclitaxel)



EVIDENCIA

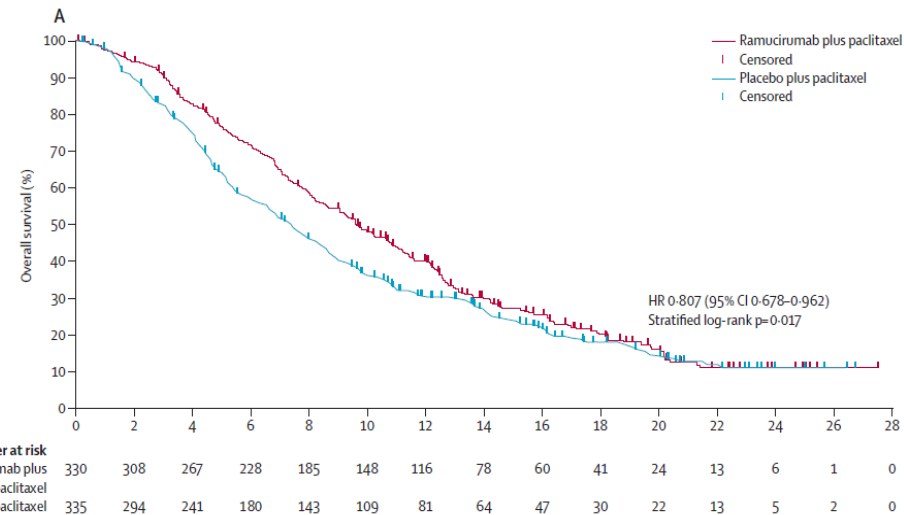
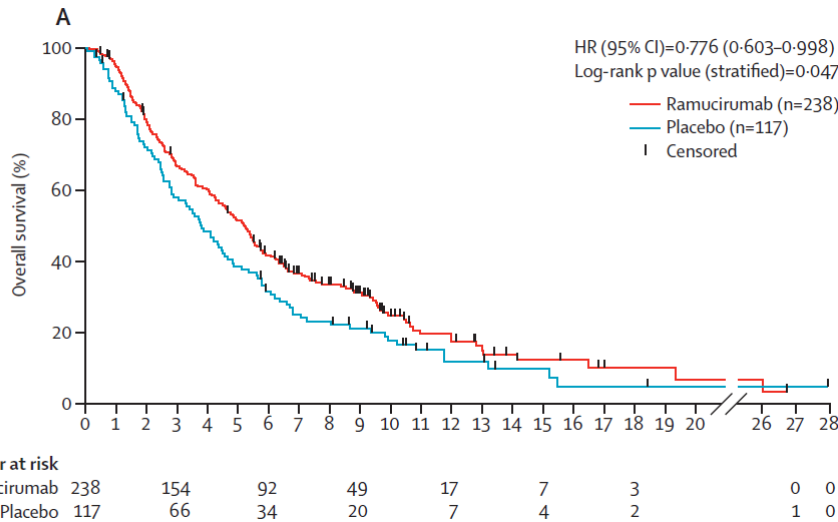
ECA de ramucirumab:

REGARD

- RAMU en monoterapia
- Fase III
- Control placebo
- Aleatorización 2:1
- SG 1,4 meses

RAINBOW

- RAMU + paclitaxel
- Fase III
- Control placebo
- Aleatorización 1:1
- SG 2,2 meses (*AUC: SG 1,57 meses*)





RAMUCIRUMAB ($\pm$ paclitaxel)

EVIDENCIA

ECA de ramucirumab:

REGARD: SG=1,4 meses

RAINBOW: SG=2,2 meses, AUC SG=1,57 meses

- **Supervivencia mínima de 3-4 meses** ⁽¹⁾ para considerar resultado significativo
 - * *Por analogía al Ca páncreas, que se comporta de forma similar al cáncer gástrico en cuanto a pronóstico y evolución)*
- Escala ESMO-MCBS 2015 ⁽²⁾ asigna a una **puntuación de 2** (escala 1-5).
 - * *Clasificación de terapias según relevancia del beneficio clínico*

Beneficio clínico reducido, alto coste

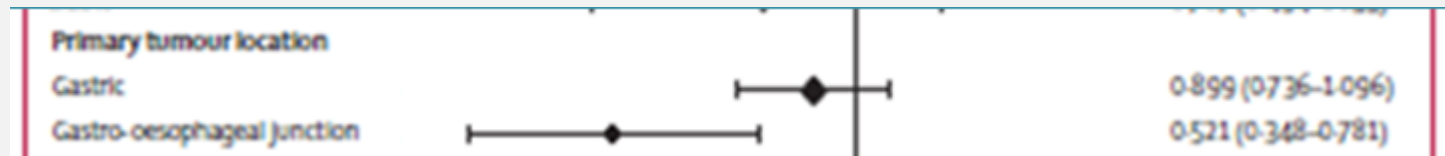
(1) Ellis LM, Bernstein DS, Voest EE, Berlin JD, Sargent D, Cortazar P et al. American Society of Clinical Oncology perspective: Raising the bar for clinical trials by defining clinically meaningful outcomes. *J Clin Oncol*. 2014 Apr 20;32(12):1277-80.

(2) Cheryn, Nathan I., et al. A standardised, generic, validated approach to stratify the magnitude of clinical benefit that can be anticipated from anti-cancer therapies: the European Society for Medical Oncology Magnitude of Clinical Benefit Scale (ESMO-MCBS). *Annals of Oncology*, 2015, p. mdv249.

RAMUCIRUMAB ($\pm$ paclitaxel)



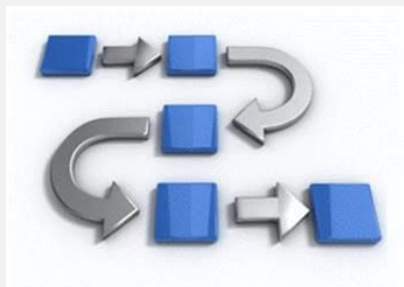
Se ha observado una posible diferencia según la localización del tumor primario en el análisis por subgrupos en la variable SG del estudio **RAINBOW**



Beneficio clínico reducido, alto coste

¿Sería razonable usar RAMU sólo en pacientes con adenocarcinoma de la unión GE?

ANÁLISIS DE SUBGRUPOS



FASES

1.- Seleccionar factores con posibles diferencias

Localización tumor
primario

2.- Valorar la similitud o fiabilidad de la diferencia

Interacción

Pre-especificación

Plausibilidad

Consistencia

3.- Recomendaciones → **CHECKLIST**

PREGUNTAS PRELIMINARES

ESTUDIO

¿Mayor nivel de evidencia?

☒

SÍ

Validez del análisis
dependerá de la propia
validez del estudio

☐

NO

Es ECA fase III

VARIABLE

¿Clara relevancia clínica?

☒

SÍ

Para la toma de
decisiones

☐

NO

Es SG

SECUENCIA TEMPORAL

¿Factor antes de intervención?

☒

SÍ

Análisis posteriores violan
principios de randomización y
tienen menor validez

☐

NO

Los pacientes presentan el
factor antes de recibir fármaco

SELECCIÓN FACTORES

¿Interacción estadística?

☒

SÍ

¿Puede deberse al
azar?

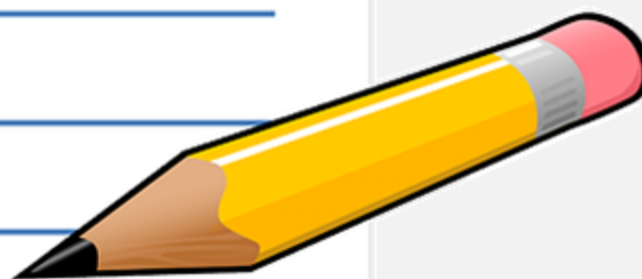
☐

NO

P= 0,0176



CHECKLIST





CRITERIO	Cuestión	Respuestas posibles	Fiabilidad del análisis	Orientaciones
I-ASOCIACIÓN ESTADÍSTICA	CUESTIONES ESENCIALES			
	1. Interacción: ¿Existe interacción, es decir, diferencia estadísticamente significativa entre el resultado de los distintos subgrupos?	Sí, con $p < 0,05$.	Probable	Se considera significativa $p < 0,05$. Pero, como al dividir en subgrupos la potencia estadística baja, se puede considerar indicativa $p < 0,1$, o incluso $p < 0,2$ en regresión multivariante, si los autores lo especifican. En muchos estudios no se aporta p de interacción. Es posible calcularla. Una p reducida o un análisis de interacción multivariante disminuyen la incertidumbre sobre la asociación, especialmente cuando se han analizado muchos factores distintos. (*) Se admitiría una excepción a la exigencia de interacción para considerar los resultados de un subgrupo: cuando por un defecto de diseño, el estudio mezcla poblaciones muy heterogéneas con respecto al efecto esperable, que justificarían estudios distintos. Podría requerirse un resultado confirmatorio en cada subgrupo.
		Sí, con $p \geq 0,05$ pero $< 0,1$ (o $p < 0,2$ en regresión multivariante si así se preespecifica)	Dudosa	
		Ninguna de las anteriores	Nula(*)	
	2. Preespecificación: ¿Estaba preespecificado en la metodología estudiar ese factor?	Sí	Probable	Es fácil encontrar <i>post hoc</i> un grupo que presente diferencias; por eso es esencial que el subgrupo estudiado estuviera preespecificado para su análisis. En su defecto, se puede considerar preespecificado si fue elegido para estratificar la aleatorización.
		No, o no se conoce	Nula	
	Cuestiones adicionales			
	3. Población (N) de los subgrupos	$N > 50$ /rama (> 100 en total)	Probable	Una N demasiado reducida facilita la descompensación de las características basales en las ramas estudiadas.
		$N < 50$ en alguna rama	Dudosa	
	4. Número de factores estudiados: contar los factores preespecificados /de estratificación (*) analizados por subgrupos en el estudio.	1-9 factores preespecificados analizados	Probable	Cuantos más factores se analicen, mayor es la probabilidad de encontrar interacción azarosa o por ramas descompensadas en factores pronósticos, sobre todo si la p no es muy reducida. (*) Si el factor es preespecificado, contar nº de factores preespecificados. Si es usado en estratificación, pero no preespecificado (caso raro) contar preespecificados + de estratificación. Si el factor no es preespecificado, asumir siempre > 10 , independientemente de los que aparezcan.
		10 o más factores	Dudosa	
	5. Resultado global del estudio	Positivo, pero podría haber diferencias entre subgrupos.	Probable	Cuando según el diseño planificado del estudio no se ha podido demostrar el beneficio, el análisis de subgrupos por sí solo tiene una fiabilidad limitada.
		Negativo, pero podría ser positivo en subgrupos	Dudosa	
	VALORACIÓN GLOBAL del criterio "ASOCIACIÓN ESTADÍSTICA"	Todos "probables"	Probable	La valoración de este criterio de "asociación estadística" se integra luego con las de "plausibilidad biológica" y "consistencia"
		1 "dudosa", 0 "nulas"	Posible	
		2 "dudosas", 0 "nulas"	Dudosa	
		≥ 3 "dudosas" o alguna "nula"	Nula	



CRITERIO	Cuestión	Respuestas posibles	Fiabilidad del análisis	Orientaciones
II- PLAUSIBILIDAD BIOLÓGICA	¿Existe una hipótesis plausible para justificar la relación de causa y efecto?	Sí, basada en literatura previa	Probable	Las diferencias deben ser explicables mediante una hipótesis aceptable, acorde al conocimiento actual de la enfermedad. Las bases para esta hipótesis deben estar presentes en la literatura. Que se haya considerado dicho factor en la estratificación muestra que se preveía su posible influencia en el resultado. Cuando un factor cuantitativo divide la población en tres o más subgrupos, resultados compatibles con un gradiente de efecto (efecto "dosis-respuesta") apoyarían la hipótesis, y al contrario.
		Es posible enunciar hipótesis razonable, pero sin base en literatura previa	Posible	
		Se puede enunciar una hipótesis dudosa y sin base en literatura previa	Dudosa	
		Sin hipótesis plausible	Nula	

CRITERIO	Cuestión	Respuestas posibles	Fiabilidad del análisis	Orientaciones
III- CONSISTENCIA	¿Existe consistencia con resultados obtenidos en otros estudios similares?	Sí, estudios similares muestran la interacción, o aparece interacción en meta-análisis y los estudios son homogéneos	Probable	Si existen estudios similares, los resultados en los grupos de pacientes estudiados deben ser compatibles con el hallazgo analizado. Estudios "similares" son aquellos con la misma intervención y control, en la misma patología. Si en dos o más se presenta la interacción y no hay resultados contradictorios, la consistencia es probable. En su ausencia de estudios similares, se pueden analizar otros con fármacos o patologías relacionadas, siempre que se estime que en ellos debería cumplirse la misma hipótesis que apoya su plausibilidad biológica. En tal caso, la consistencia máxima sería "posible". Si no hay otros estudios con los que comparar, se puede recurrir a comprobar si existe falta de consistencia interna con los resultados de otras variables relevantes relacionadas con la estudiada. Está dificultado, por lo reducido del tamaño de la muestra que aumenta el error beta, que en todos los estudios aparezca interacción significativa. Puede recurrirse a metaanálisis de resultados por subgrupos y, una vez observada la interacción en el análisis global, comprobar su homogeneidad, que apoyaría la diferencia (probable).
		No, pero existen resultados compatibles (sin interacción en todos, pero sugestivos de posibles diferencias en el mismo sentido).	Posible	
		No existen otros estudios con los que comparar. No existe, al menos, falta de consistencia interna.	Dudosa	
		No se cumple ninguna de las condiciones anteriores	Nula	

RAMUCIRUMAB

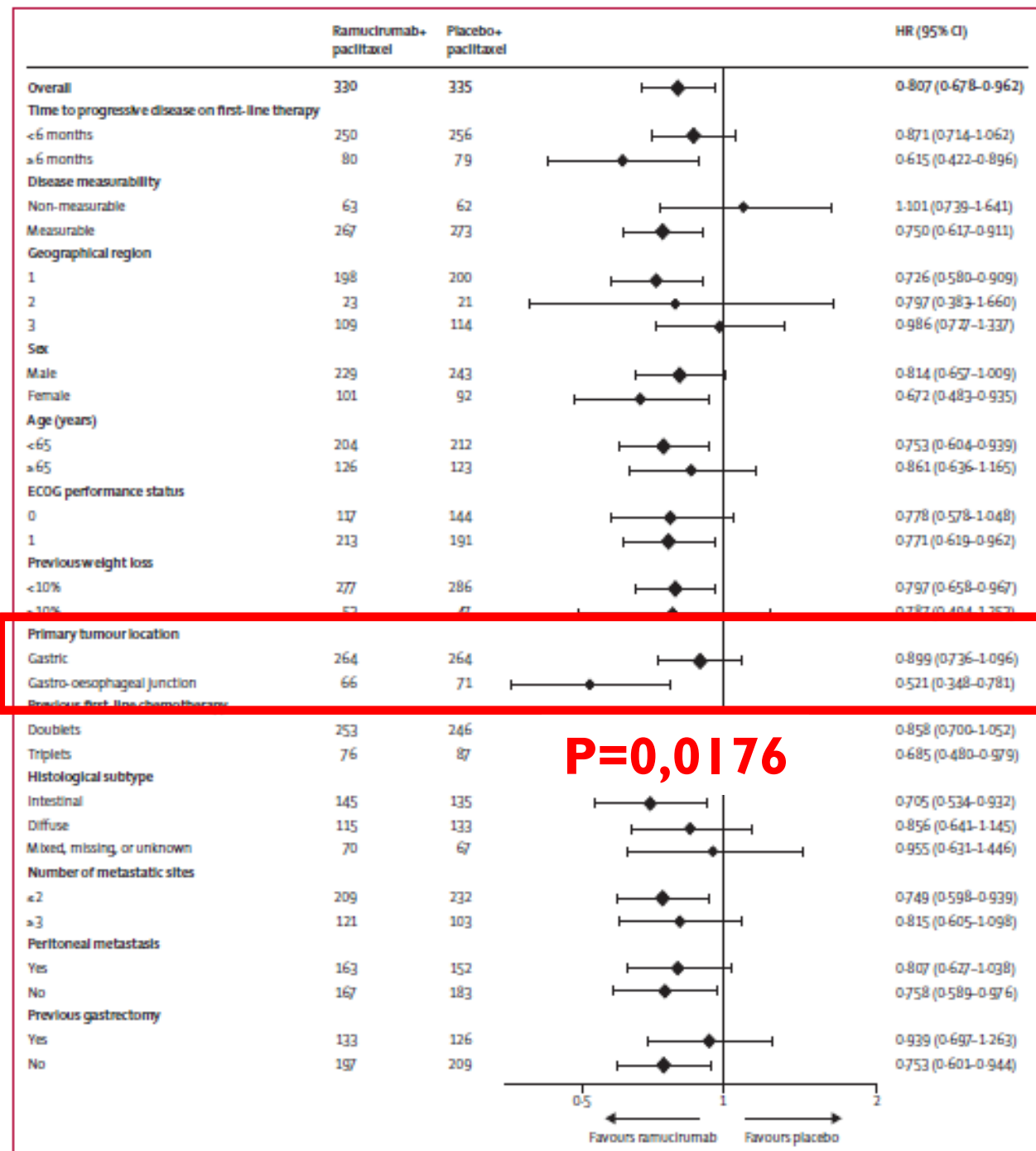
Estudio RAINBOW

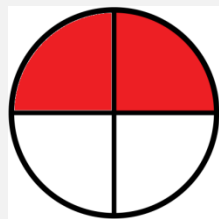
Figure 3: Forest plot for subgroup univariate analyses of overall survival

Data are stratified HR (95% CI).

The size of the diamonds is proportional to the size of the subgroup.

Geographic regions are defined as region 1: Europe, Israel, Australia, and the USA; region 2: Argentina, Brazil, Chile, and Mexico; and region 3: Japan, South Korea, Hong Kong, Singapore, and Taiwan.





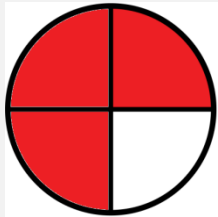
INTERACCIÓN y PREESPECIFICACIÓN

P < 0,05

**Pac/ram
a > 100**

**Total 17
factores**

CRITERIO	Cuestión	Respuestas posibles	Fiabilidad del análisis	Orientaciones
I-ASOCIACIÓN ESTADÍSTICA P < 0,05 Pac/ram a > 100 Total 17 factores	CUESTIONES ESENCIALES			
	1. Interacción: ¿Existe interacción, es decir, diferencia estadísticamente significativa entre el resultado de los distintos subgrupos?	Sí, con $p < 0,05$.	Probable	Se considera significativa $p < 0,05$. Pero, como al dividir en subgrupos la potencia estadística baja, se puede considerar indicativa $p < 0,1$, o incluso $p < 0,2$ en regresión multivariante, si los autores lo especifican. En muchos estudios no se aporta p de interacción. Es posible calcularla. Una p reducida o un análisis de interacción multivariante disminuyen la incertidumbre sobre la asociación, especialmente cuando se han analizado muchos factores distintos. (*) Se admitiría una excepción a la exigencia de interacción para considerar los resultados de un subgrupo: cuando por un defecto de diseño, el estudio mezcla poblaciones muy heterogéneas con respecto al efecto esperable, que justificarían estudios distintos. Podría requerirse un resultado confirmatorio en cada subgrupo.
		Sí, con $p \geq 0,05$ pero $< 0,1$ (o $p < 0,2$ en regresión multivariante si así se preespecifica)	Dudosa	
		Ninguna de las anteriores	Nula(*)	
	2. Preespecificación: ¿Estaba preespecificado en la metodología estudiar ese factor?	Sí	Probable	Es fácil encontrar <i>post hoc</i> un grupo que presente diferencias; por eso es esencial que el subgrupo estudiado estuviera preespecificado para su análisis. En su defecto, se puede considerar preespecificado si fue elegido para estratificar la aleatorización.
		No, o no se conoce	Nula	
	Cuestiones adicionales			
	3. Población (N) de los subgrupos	N > 50/rama (>100 en total)	Probable	Una N demasiado reducida facilita la descompensación de las características basales en las ramas estudiadas.
		N < 50 en alguna rama	Dudosa	
	4. Número de factores estudiados: contar los factores preespecificados /de estratificación (*) analizados por subgrupos en el estudio.	1-9 factores preespecificados analizados	Probable	Cuanto más factores se analicen, mayor es la probabilidad de encontrar interacción azarosa o por ramas descompensadas en factores pronósticos, sobre todo si la p no es muy reducida. (*) Si el factor es preespecificado, contar nº de factores preespecificados. Si es usado en estratificación pero no preespecificado (caso raro) contar preespecificados + de estratificación. Si el factor no es preespecificado, asumir siempre >10, independientemente de los que aparezcan.
		10 o más factores	Dudosa	
	5. Resultado global del estudio	Positivo, pero podría haber diferencias entre subgrupos.	Probable	Cuando según el diseño planificado del estudio no se ha podido demostrar el beneficio, el análisis de subgrupos por sí solo tiene una fiabilidad limitada.
		Negativo, pero podría ser positivo en subgrupos	Dudosa	
	VALORACIÓN GLOBAL del criterio "ASOCIACIÓN ESTADÍSTICA"	Todos "probables"	Probable	La valoración de este criterio de "asociación estadística" se integra luego con las de "plausibilidad biológica" y "consistencia"
		1 "dudosa", 0 "nulas"	Posible	
		2 "dudosas", 0 "nulas"	Dudosa	



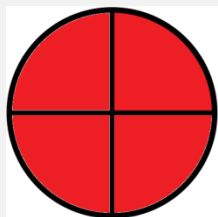
PLAUSIBILIDAD BIOLÓGICA



CRITERIO	Cuestión	Respuestas posibles	Fiabilidad del análisis	Orientaciones
II- PLAUSIBILIDAD BIOLÓGICA	¿Existe una hipótesis plausible para justificar la relación de causa y efecto?	Sí, basada en literatura previa	Probable	Las diferencias deben ser explicables mediante una hipótesis aceptable, acorde al conocimiento actual de la enfermedad. Las bases para esta hipótesis deben estar presentes en la literatura. Que se haya considerado dicho factor en la estratificación muestra que se preveía su posible influencia en el resultado.
		Es posible enunciar hipótesis razonable, pero sin base en literatura previa	Posible	
		Se puede enunciar una hipótesis dudosa y sin base en literatura previa	Dudosa	Cuando un factor cuantitativo divide la población en tres o más subgrupos, resultados compatibles con un gradiente de efecto (efecto "dosis-respuesta") apoyarían la hipótesis, y al contrario.
		Sin hipótesis plausible	Nula	

Se podría enunciar hipótesis que relacionase
SG con localización tumor primario...

ESCASO APOYO EN LA LITERATURA



CONSISTENCIA

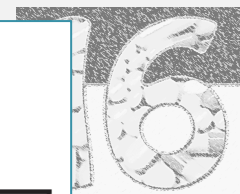
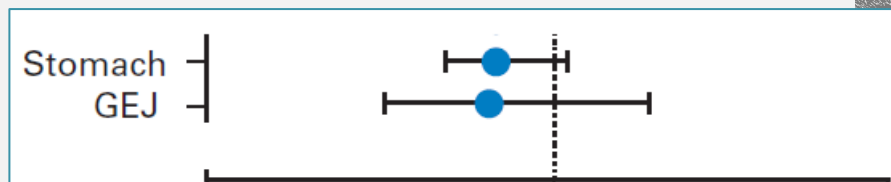


CRITERIO	Cuestión	Respuestas posibles	Fiabilidad del análisis	Orientaciones
III- CONSISTENCIA	¿Existe consistencia con resultados obtenidos en otros estudios similares?	Si, estudios similares muestran la interacción, o aparece interacción en meta-análisis y los estudios son homogéneos	Probable	Si existen estudios similares, los resultados en los grupos de pacientes estudiados deben ser compatibles con el hallazgo analizado. Estudios "similares" son aquellos con la misma intervención y control, en la misma patología. Si en dos o más se presenta la interacción y no hay resultados contradictorios, la consistencia es probable. En su ausencia de estudios similares, se pueden analizar otros con fármacos o patologías relacionadas, siempre que se estime que en ellos debería cumplirse la misma hipótesis que apoya su plausibilidad biológica. En tal caso, la consistencia máxima sería "posible". Si no hay otros estudios con los que comparar, se puede recurrir a comprobar si existe falta de consistencia interna con los resultados de otras variables relevantes relacionadas con la estudiada. Está dificultado, por lo reducido del tamaño de la muestra que aumenta el error beta, que en todos los estudios aparezca interacción significativa. Puede recurrirse a metaanálisis de resultados por subgrupos y, una vez observada la interacción en el análisis global, comprobar su homogeneidad, que apoyaría la diferencia (probable).
		No, pero existen resultados compatibles (sin interacción en todos, pero en el mismo sentido).	Posible	
		No existen otros estudios con los que comparar. No existe, al menos, falta de consistencia interna.	Dudosa	
		No se cumple ninguna de las condiciones anteriores	Nula	

REGARD: no interacción

AVAGAST: no interacción

	Ramucirumab group	Placebo group		HR (95% CI)
Primary tumour				
Gastric	179	85		0.823 (0.608-1.114)
Gastro-oesophageal junction	59	32		0.756 (0.472-1.211)



Resultados checklist



<u>CRITERIO</u>	<u>VALORACIÓN</u> probable, posible, dudosa o nula	<u>PUNTUACIÓN</u> Probable: 3 puntos. Posible: 2 puntos. Dudosa: 0 puntos. Nula: -3 puntos.
ASOCIACIÓN ESTADÍSTICA	Posible	2
PLAUSIBILIDAD BIOLÓGICA	Posible	2
CONSISTENCIA	Nula	-3
APLICABILIDAD PRÁCTICA (SUMA): 7-9 probable, 5-6 posible, 3-4 dudosa, <3 nula		1

A) RECOMENDACIONES PARA CONSIDERAR EL RESULTADO DE UN SUBGRUPO EN LA TOMA DE DECISIONES CLÍNICAS

Aplicabilidad	Recomendación
PROBABLE	Los resultados del análisis de subgrupos deberían ser tenidos en cuenta en la práctica clínica, en tanto no se disponga de un ensayo clínico diseñado específicamente para confirmar o descartar la diferencia de efecto.
POSIBLE	Los resultados del análisis de subgrupos podrían ser aplicables a la práctica clínica con prudencia, si la situación lo aconseja. Deberían ser tenidos en consideración si la intervención resulta más tóxica, difícil de utilizar o muy cara con respecto a sus alternativas.
DUDOSA	El análisis de subgrupos no debe ser tenido en consideración para la práctica clínica. * Excepcionalmente, podría sopesarse su aplicación en el caso de una intervención mucho más tóxica, muy difícil de usar o de coste inasumible, que aconsejase restringir el uso a situaciones de beneficio indudable.
NULA	El análisis de subgrupos no muestra diferencias que puedan ser consideradas de forma fiable en la práctica clínica. Los resultados por subgrupos no deben usarse para valorar la eficacia: el resultado atribuible a cada subgrupo no es otro que el global del estudio.