



LECTURA CRÍTICA DE ENSAYOS CLÍNICOS

--

PUNTOS CLAVE

Silvia Fénix Caballero
H. U. Puerto Real
GENESIS-SEFH
24/11/2022



VALIDEZ INTERNA

- Aleatorización y comparador
- Pregunta PICO
- Seguimiento
- Ciego
- Diseño

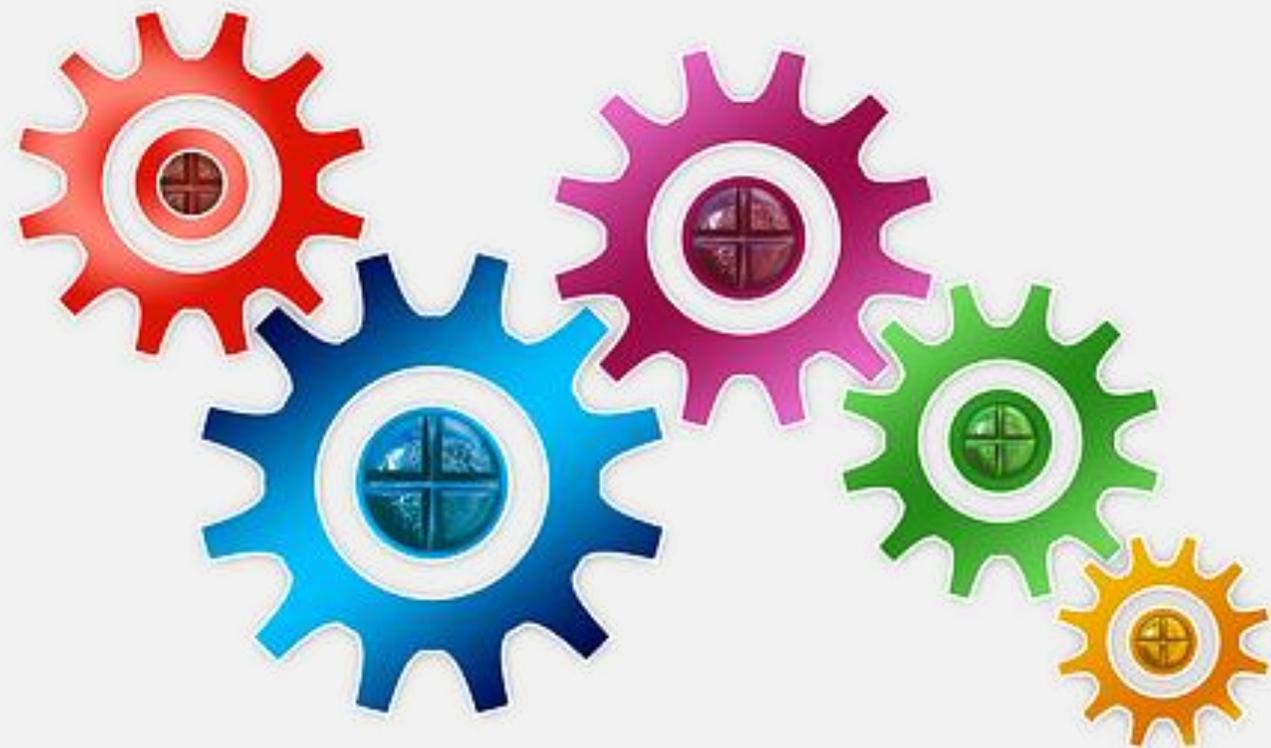
RESULTADOS

- Relevancia clínica variables
- Significación estadística vs relevancia clínica
- Análisis de subgrupos
- Valoración de la precisión

VALIDEZ EXTERNA

- Similitud pacientes del ECA con los de la práctica clínica

Validez
interna



HERRAMIENTA VALIDADA DE EVALUACIÓN CALIDAD METODOLÓGICA

Escala de Jadad / sistema de puntuación de calidad de Oxford

Criterios	Puntuación
¿Se describe el estudio como aleatorizado?(*)	
¿Se describe el estudio como doble ciego?(*)	
¿Se describen las pérdidas y retiradas del estudio?(*)	
¿Es adecuado el método de aleatorización?(**)	
¿Es adecuado el método de doble ciego?(**)	
(*) Sí= 1 / No= 0	
(**) Sí= 1 / No= -1	

Puntúa los EC de 0 a 5

Si puntuación < 3, el EC se considera de baja calidad



Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJM, Gavaghan DJ, McQuay HJ. Assessing the quality of reports of randomized clinical trials: is blinding necessary?. *Controlled Clinical Trials*. 1996; 17(1):1-12.

E.C. NO CONTROLADO



¡¡SESGO!!

MAYOR ESTIMACIÓN DEL EFECTO

E.C. CONTROLADO



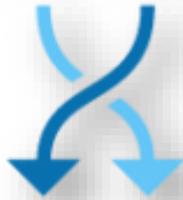
BENEFICIO
CLÍNICO

RESULTADO CLÍNICO

EVOLUCIÓN NATURAL ENFERMEDAD
EXPECTATIVAS PACIENTE/INVESTIGADOR



GRUPO DE ESTUDIO



ALEATORIZACIÓN



GRUPO DE
TRATAMIENTO



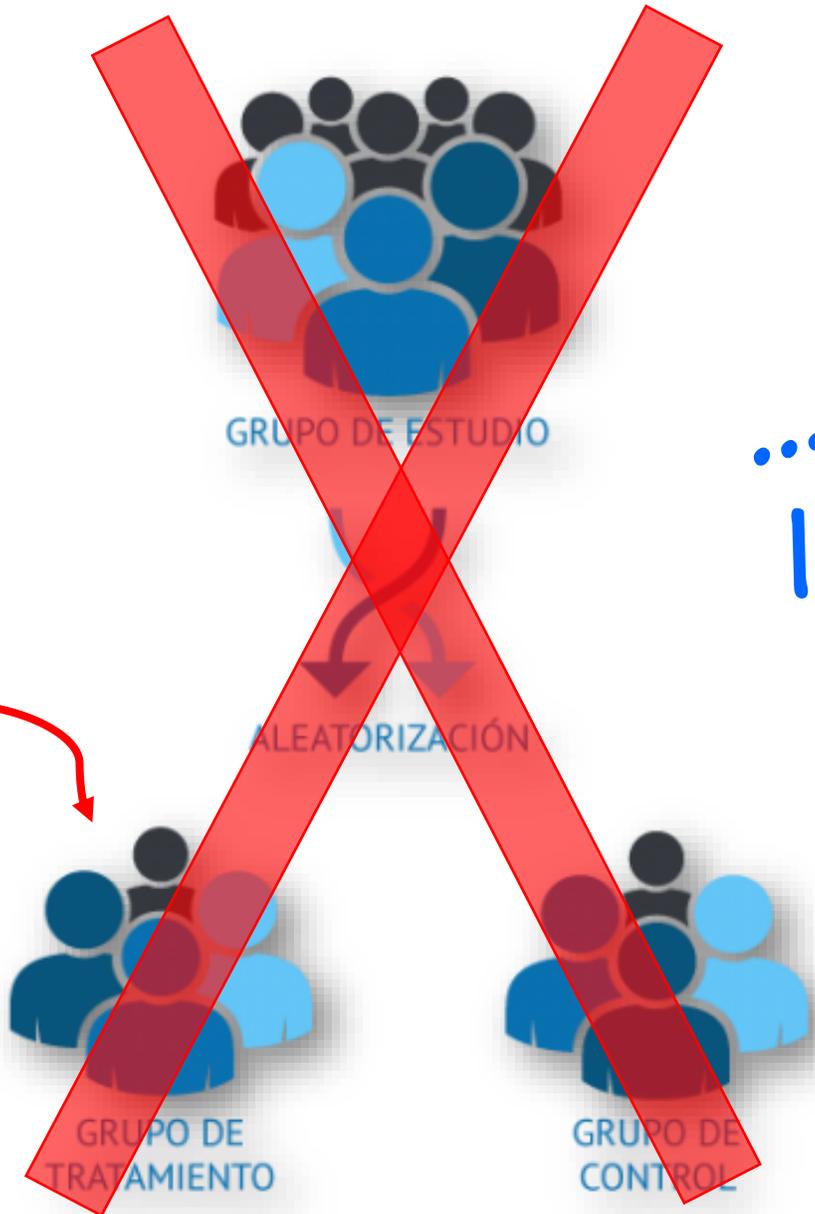
GRUPO DE
CONTROL

ENSAYOS CLÍNICOS ALEATORIZADOS

★ Participantes desconocen secuencia asignación



★ Impide asignar de forma sesgada



...¿Qué ocurre con
los

ENSAYOS CLÍNICOS **NO** ALEATORIZADOS?

- HOMOGENEIDAD
- GRUPOS EQUILIBRADOS
- SESGO SELECCIÓN**



PREGUNTA PICO

Paciente

- ¿Cómo se describe la población estudiada?

Intervención

- ¿Qué se está evaluando?

Comparación

- ¿Alternativa con la que se compara?

Outcomes

- ¿Qué beneficio clínico medible se espera?

IMPORTANTE: **definición** y **medida** de variable

DIAGO JORGE DIAZ NAVARRO



19%

FINALIZAN PREMATURAMENTE

Reclutamiento factor limitante

**RECLUTAMIENTO
COMPLETO**



89%

SE RETRASAN

- X** Nº pacientes esperado
- X** Nº eventos esperados
- ?** Algún brazo es claramente perjudicial o beneficioso

Alteraciones
protocolo
previsto

Afecta + ó -

Poder
estadístico del
ensayo

Detección diferencias
estadísticamente
significativas

Tamaño muestral planificado
Nº análisis intermedios
Reglas de parada

SEGUIMIENTO

PÉRDIDAS POST ALEATORIZACIÓN



**ABANDONOS
(DROP OUT)**



**RETIRADAS
(WITHDRAWAL)**

*-nuevos criterios de exclusión
-RAM graves ...*

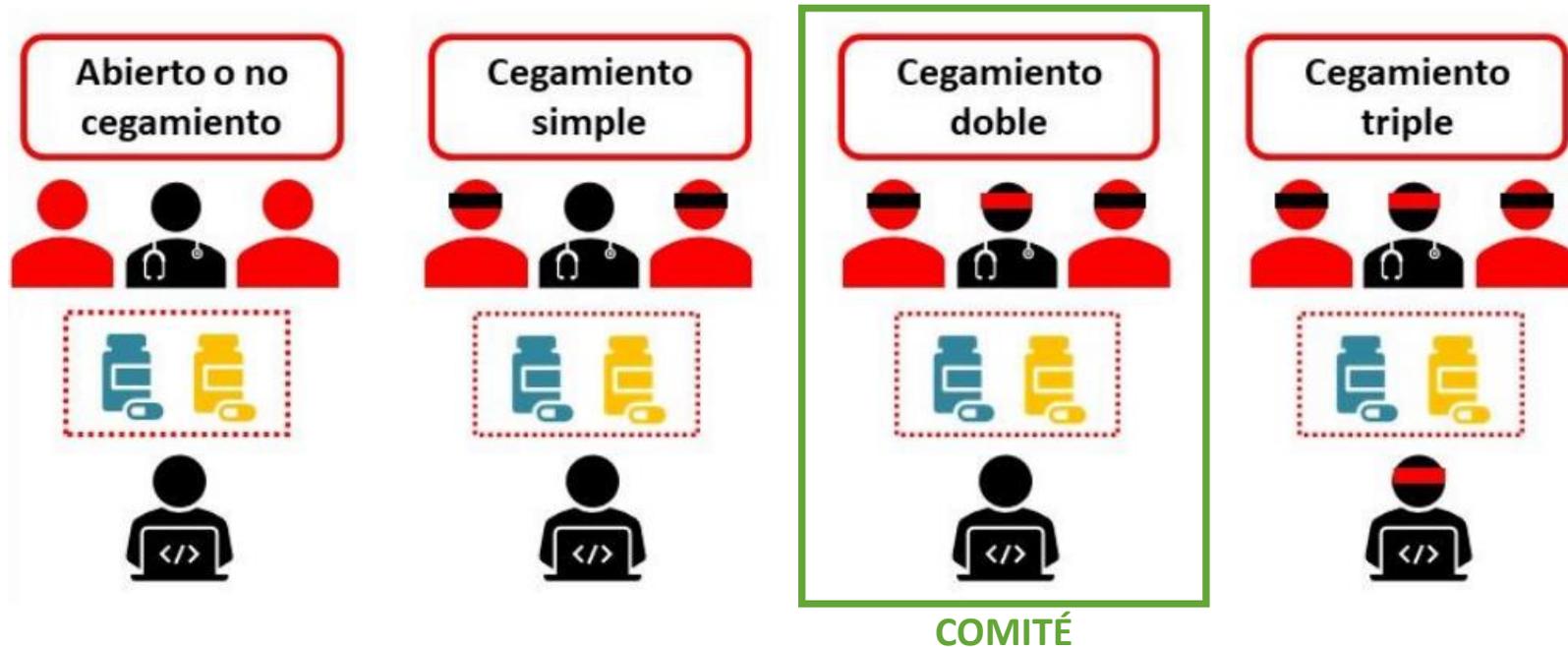


**PÉRDIDAS DE
SEGUIMIENTO**

no realizan controles

CIEGO

Desconocimiento de la intervención que reciben los grupos estudiados



La relevancia del «ciego» en la validez

TIPO DE VARIABLE

- Subjetivas (dolor, calidad de vida)
- Objetivas (mortalidad, fractura ósea)



GRUPOS BALANCEADOS

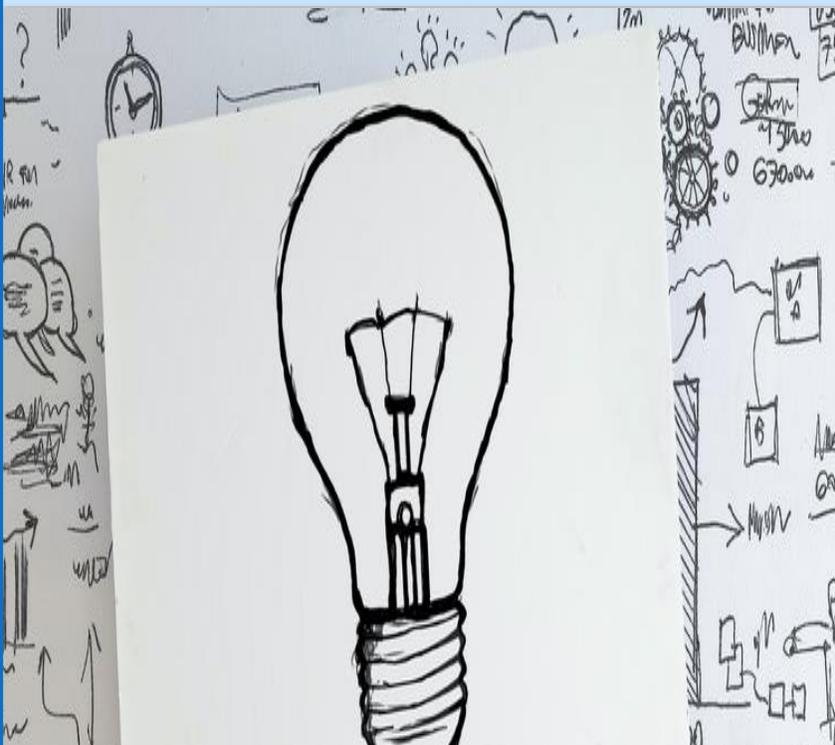
- ⇒ **Características basales balanceadas** → inferencias estadísticas y de causalidad, diferencias debidas al tto
- ⇒ **Desbalance**: asignar cierto tipo de pacientes al grupo experimental o control: RIESGO
- ⇒ La **aleatorización** ↓ sesgo de desbalance



Resultados



Relevancia clínica de las variables



ORIENTADAS A LA ENFERMEDAD



- Intermedias / subrogadas / subclínicas
- Parámetros de laborat (habitualmente)
- Utilidad según capacidad predictora

Niveles colesterol
Retraso parto pretérmino
Densidad mineral ósea
SLP

ORIENTADAS AL PACIENTE

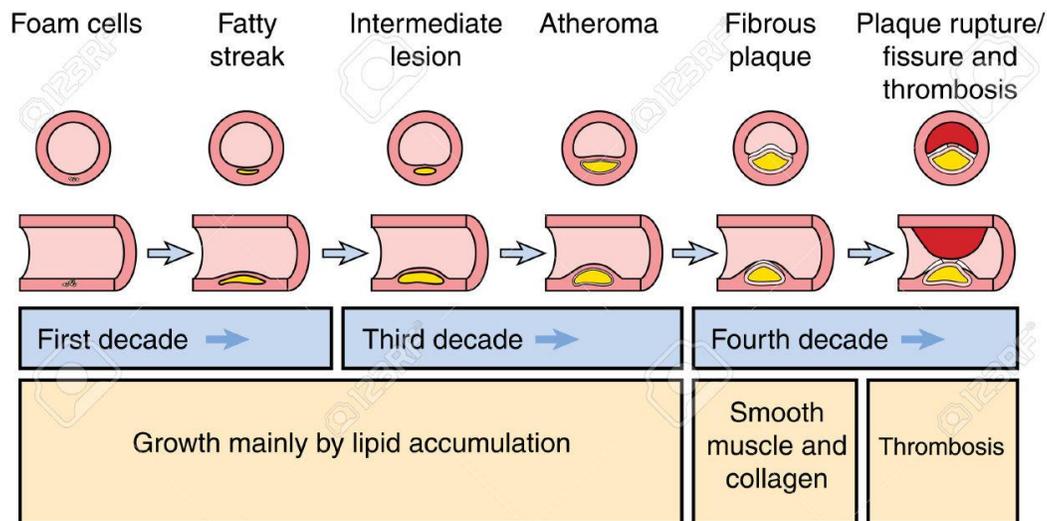


- Duras / finales / clínicas
- Eventos clínicos importantes
- Objetividad determinación

IAM
Nacimiento niño sano
Fractura de cadera
SG

SUBROGADA: paso fisiopatológico del desarrollo de la enfermedad

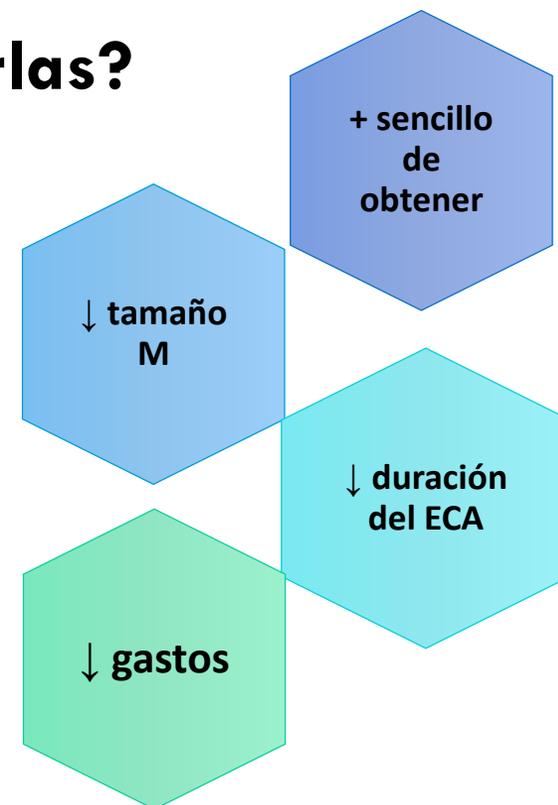
↑ C-LDL



VARIABLE SUBROGADA / INTERMEDIA

Utilidad → depende de su capacidad predictora

¿Por qué usarlas?





Significación estadística

- Implica utilizar términos comparativos de dos hipótesis

H_0 (hipótesis nula)



H_1 (hipótesis alternativa)



"p": probabilidad de que exista diferencia (aceptar H1)

Valor 0,05





Relevancia clínica

No cálculo matemático → **JUICIO CLÍNICO**

magnitud de la
diferencia

gravedad de la
variable

Herramienta objetiva → sentido clínico a la significación estadística

Cálculo tamaño muestral

Adalimumab en AR

Statistical analysis. The power calculation was based on both the predicted ACR20 response rate and radiographic findings. A sample size of 200 patients per treatment group was estimated to provide $>95\%$ power for detecting a difference of $\geq 20\%$ in ACR20 response rates at week 24 between the placebo group and the adalimumab groups at a significance level of $\alpha = 0.05$, assuming a placebo response rate of 35%.



ESMO-MCBS EVALUATION FORMS

Evaluation form 1  For new approaches to adjuvant therapy or new potentially curative therapies

Evaluation form 2a For therapies that are not likely to be curative with primary endpoint of OS with separate sheets for:

- IF median OS with the standard treatment is ≤ 12 months 
- IF median OS with the standard treatment is >12 months, ≤ 24 months 
- IF median OS with the standard treatment is >24 months 

Evaluation form 2b For therapies that are not likely to be curative with primary endpoint PFS with separate sheets for:

- IF median PFS with standard treatment is ≤ 6 months 
- IF median PFS with standard treatment is >6 months 

Evaluation form 2c For therapies that are not likely to be curative with primary endpoint other than OS or PFS or equivalent (non-inferiority) studies

Evaluation form 3 For single-arm studies in “orphan diseases” and for diseases with “high unmet need” when primary outcome is PFS or ORR.

ESMO-MAGNITUDE OF CLINICAL BENEFIT SCALE V1.1 EVALUATION FORM 2A

For therapies that are not likely to be curative with primary endpoint of OS

If median OS with the standard treatment is ≤ 12 months

GRADE 4	HR ≤ 0.65 AND gain ≥ 3 months	<input type="radio"/>
	Increase in 2 year survival $\geq 10\%$	<input type="radio"/>
GRADE 3	HR ≤ 0.65 AND gain ≥ 2.0 - <3 months	<input type="radio"/>
GRADE 2	HR ≤ 0.65 AND gain ≥ 1.5 - <2.0	<input type="radio"/>
	HR >0.65 - 0.70 AND gain ≥ 1.5 months	<input type="radio"/>
GRADE 1	HR >0.70 OR gain <1.5 months	<input type="radio"/>

Mark with \surd if relevant

Preliminary magnitude of clinical benefit grade (highest grade scored)	4	3	2	1
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Non-curative setting grading - 5 and 4 indicates a substantial magnitude of clinical benefit

Quality of life/Grade 3-4 toxicities* assessment

Does secondary endpoint QoL show improvement?

Are there statistically significantly less grade 3-4 toxicities impacting on daily well-being?*

*This does not include alopecia, myelosuppression, but rather chronic nausea, diarrhoea, fatigue, etc.

Mark with ✓ if relevant

Adjustments

01. Upgrade 1 level if improved QoL and/or less grade 3-4 toxicities impacting daily well-being are shown
02. If there is a long term plateau in the survival curve, and OS advantage continues to be observed at 5 years, also score according to form 1 (treatments with curative potential) and present both scores i.e. A/4.

Final adjusted magnitude of clinical benefit grade

5

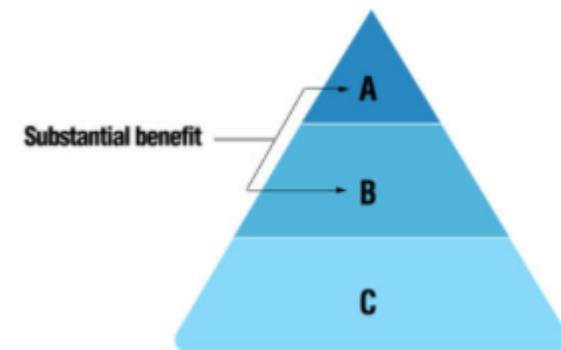
4

3

2

1

Non-curative setting grading - 5 and 4 indicates a substantial magnitude of clinical benefit



Escenario curativo



Escenario no-curativo

Filter by

Agent **Tumour** **Score** × reset selection

Tested agent(s) ▼ Tumour type ▼ Curative setting ▼

Tumour sub-type ▼ Non-curative setting ▼

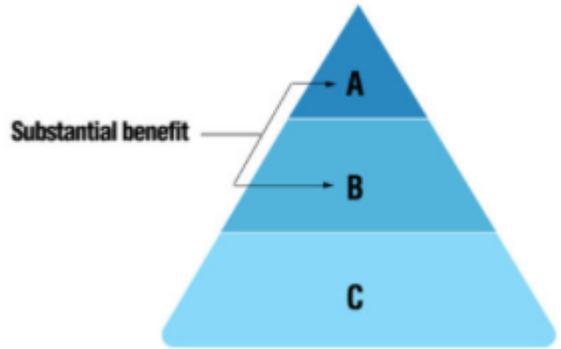
Tumour sub-group ▼

ESMO-MCBS V1.1 – Evaluation Forms

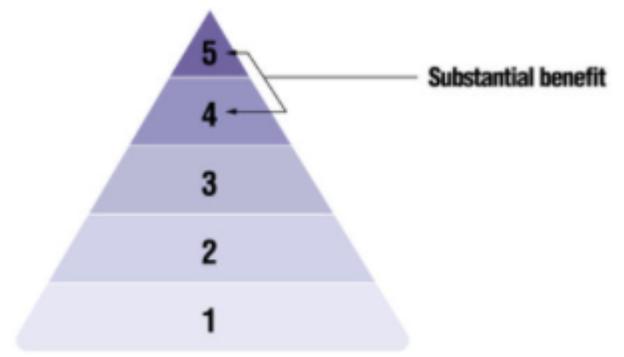
Evaluation form 1 ▼
 Evaluation form 2a ▼
 Evaluation form 2b ▼
 Evaluation form 2c ▼
 Evaluation form 3 ▼

Column view Table view

✕	Tested Agent(s)	Combined Agent(s)	Control Arm	Treatment Setting	Tumour Type	Tumour Sub-type	Tumour Sub-group	Trial Name	Ref.	Score	Scorecard
<input type="checkbox"/>	Trastuzumab emtansine (T-DM1)	-	Trastuzumab	Adjuvant treatment of patients with HER2-positive early breast cancer who have residual invasive disease in the breast and/or lymph nodes after neoadjuvant taxane-based and HER2-targeted therapy	Breast Cancer	Early Breast Cancer	HER2+	KATHERINE		A	→



Escenario curativo



Escenario no-curativo



American Society of Clinical Oncology Perspective: Raising the Bar for Clinical Trials by Defining Clinically Meaningful Outcomes

Meaningful Outcomes in Cancer Trials

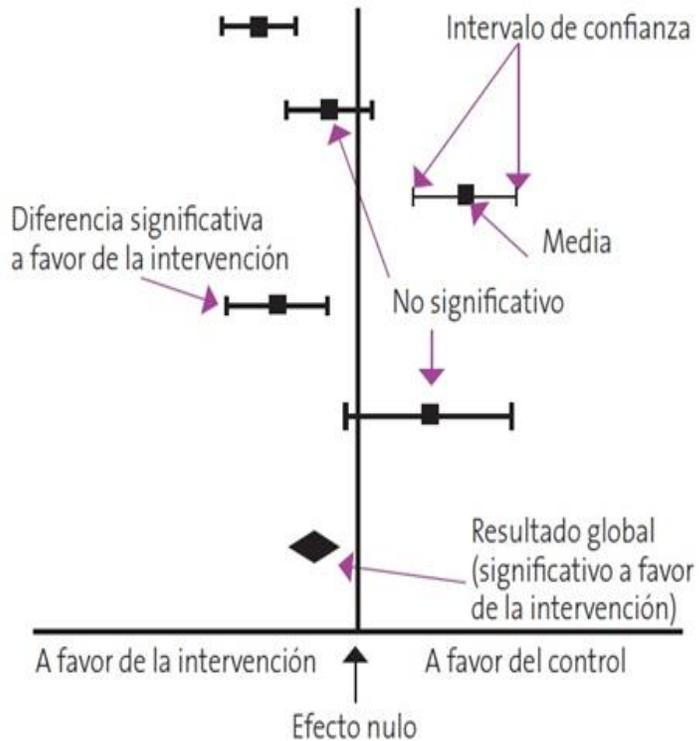
Table 1. Summary of Recommended Targets for Meaningful Clinical Trial Goals

Cancer Type	Patient Population	Current Baseline Median OS (months)	Primary End Point		Secondary End Point	
			Improvement Over Current OS That Would Be Clinically Meaningful (months)	Target HRs	Improvement in 1-Year Survival Rate (%)*	Improvement in PFS (months)
Pancreatic cancer	FOLFIRINOX-eligible patients	10 to 11 ¹⁹	4 to 5	0.67 to 0.69	48 → 63	4 to 5
Pancreatic cancer	Gemcitabine or gemcitabine/nab-paclitaxel-eligible patients	8 to 9 ^{20,21}	3 to 4	0.6 to 0.75	35 → 50	3 to 4
Lung cancer	Nonsquamous cell carcinoma	13 ²²	3.25 to 4	0.76 to 0.8	53 → 61	4
Lung cancer	Squamous cell carcinoma	10 ²³	2.5 to 3	0.77 to 0.8	44 → 53	3
Breast cancer	Metastatic triple negative, previously untreated for metastatic disease	18 ^{24,25}	4.5 to 6	0.75 to 0.8	63 → 71	4
Colon cancer	Disease progression with all prior therapies (or not a candidate for standard second- or third-line options)	4 to 6 ²⁶	3 to 5	0.67 to 0.67	25 → 35	3 to 5

Abbreviations: FOLFIRINOX, leucovorin, fluorouracil, irinotecan, and oxaliplatin; HR, hazard ratio; OS, overall survival; PFS, progression-free survival.

*Current → target.

Análisis de subgrupos



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.

¿ Privar el uso de aspirina en mujeres con enfermedad cerebrovascular ?

FALSOS -

FALSOS +





FALSOS -

FALSOS +

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

Ocurre cuando :

- ✧ N pequeña
- ✧ Diferencia del Efecto ↓

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

Ocurre cuando:

- ✧ ↑ nº de subgrupos ↑ la probabilidad de encontrar diferencias por azar

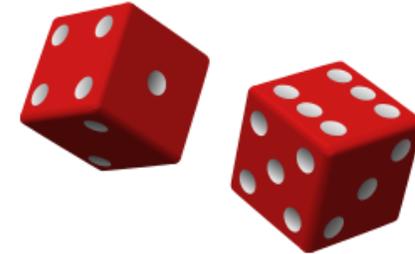
INTERACCIÓN

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

INTERACCIÓN = SIGNIFICACIÓN

Generalmente

$p < 0,05$



Subgrupos: ↓ muestra → ↓ poder estadístico

No detección de diferencias que sí existen

$P < 0,1$

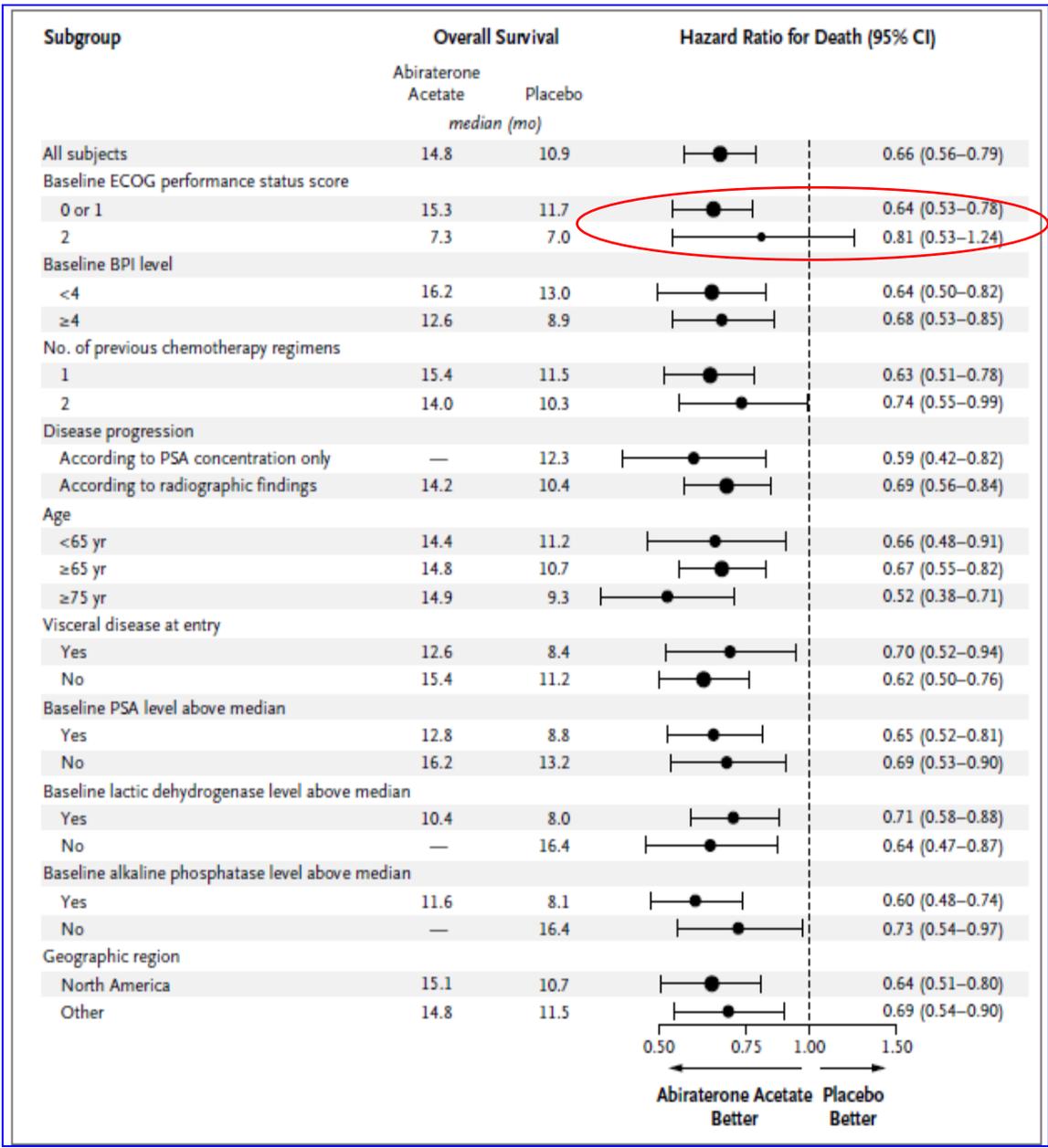
FOREST-PLOT

“DIFERENCIA INTERGRUPAL”

$p < 0,05$

$p < 0,1$

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



Muchos análisis de subgrupos (el 47%) carecen del test de interacción.

Gabler et al., Trials. 2009;10:43.

Calculadoras



$p < 0,05$

$p < 0,1$

CALCULADORA DE INTERACCIÓN

Joaquín Primo -

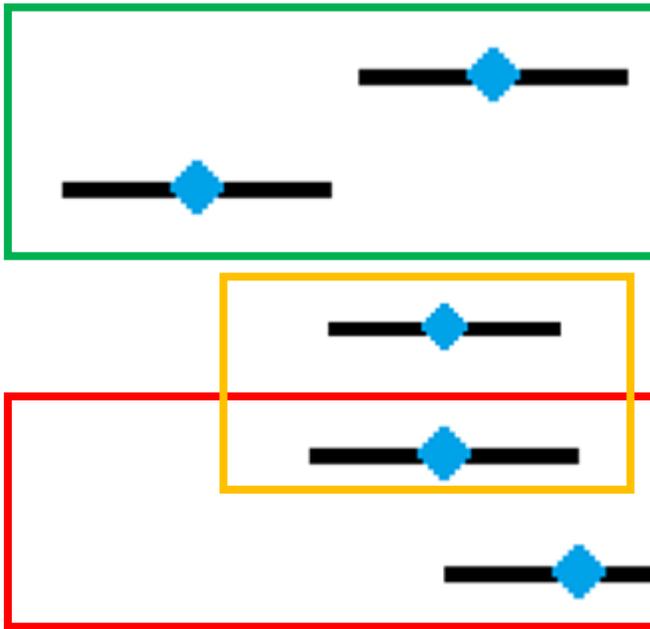
2 subgrupos

	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	Razón de RR, OR o HR	0,76
Error estándar	0,1929	0,1070	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 1 | Medias 2 | Proporciones 1 | Proporciones 2 | **RR-OR-HR** | + | :

GRÁFICAMENTE



Subgroup	Overall Survival		Hazard Ratio for Death (95% CI)	
	Abiraterone Acetate	Placebo	HR	95% CI
All subjects	14.8	10.9	0.66	(0.56–0.79)
Baseline ECOG performance status score				
0 or 1	15.3	11.7	0.64	(0.53–0.78)
2	7.3	7.0	0.81	(0.53–1.24)
Baseline BPI level				
<4	16.2	13.0	0.64	(0.50–0.82)
≥4	12.6	8.9	0.68	(0.53–0.85)
No. of previous chemotherapy regimens				
1	15.4	11.5	0.63	(0.51–0.78)
2	14.0	10.3	0.74	(0.55–0.99)
Disease progression				
According to PSA concentration only	—	12.3	0.59	(0.42–0.82)
According to radiographic findings	14.2	10.4	0.69	(0.56–0.84)
Age				
<65 yr	14.4	11.2	0.66	(0.48–0.91)
≥65 yr	14.8	10.7	0.67	(0.55–0.82)
≥75 yr	14.9	9.3	0.52	(0.38–0.71)
Visceral disease at entry				
Yes	12.6	8.4	0.70	(0.52–0.94)
No	15.4	11.2	0.62	(0.50–0.76)
Baseline PSA level above median				
Yes	12.8	8.8	0.65	(0.52–0.81)
No	16.2	13.2	0.69	(0.53–0.90)
Baseline lactic dehydrogenase level above median				
Yes	10.4	8.0	0.71	(0.58–0.88)
No	—	16.4	0.64	(0.47–0.87)
Baseline alkaline phosphatase level above median				
Yes	11.6	8.1	0.60	(0.48–0.74)
No	—	16.4	0.73	(0.54–0.97)
Geographic region				
North America	15.1	10.7	0.64	(0.51–0.80)
Other	14.8	11.5	0.69	(0.54–0.90)

nar



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**

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**

Ej: signos del Zodíaco en el ensayo ISIS-2

En ISIS-2 aspirina era ineficaz en pacientes de signo del zodiaco **Libra y Géminis**, y beneficioso en el resto

Aspirina vs. placebo en IAM

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

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

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

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

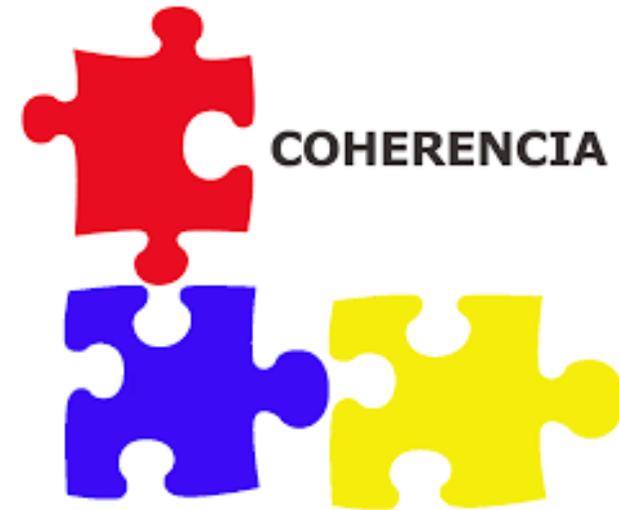
Esa amarga lamentación debieron exclamar los nacidos bajo el signo zodiacal de géminis que leyeron el análisis de subgrupos del estudio ISIS-2 que demostraba ausencia de efecto de la aspirina frente a placebo tras un infarto agudo de miocardio, en los pacientes pertenecientes a los signos zodiacales 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? Aparentemente la medicina personalizada no llega



CONSISTENCIA

Comportamiento en

- Otras variables
- Otras indicaciones
- Otros fármacos (mecanismo acción)



Valoración de la precisión

- ➔ Ensayos clínicos → estimación efecto en pacientes concretos
- ➔ Estimación del efecto en la población general

Conocer

Magnitud efecto
Precisión resultados → intervalos de confianza



Verdadera magnitud del efecto con un grado prefijado de seguridad



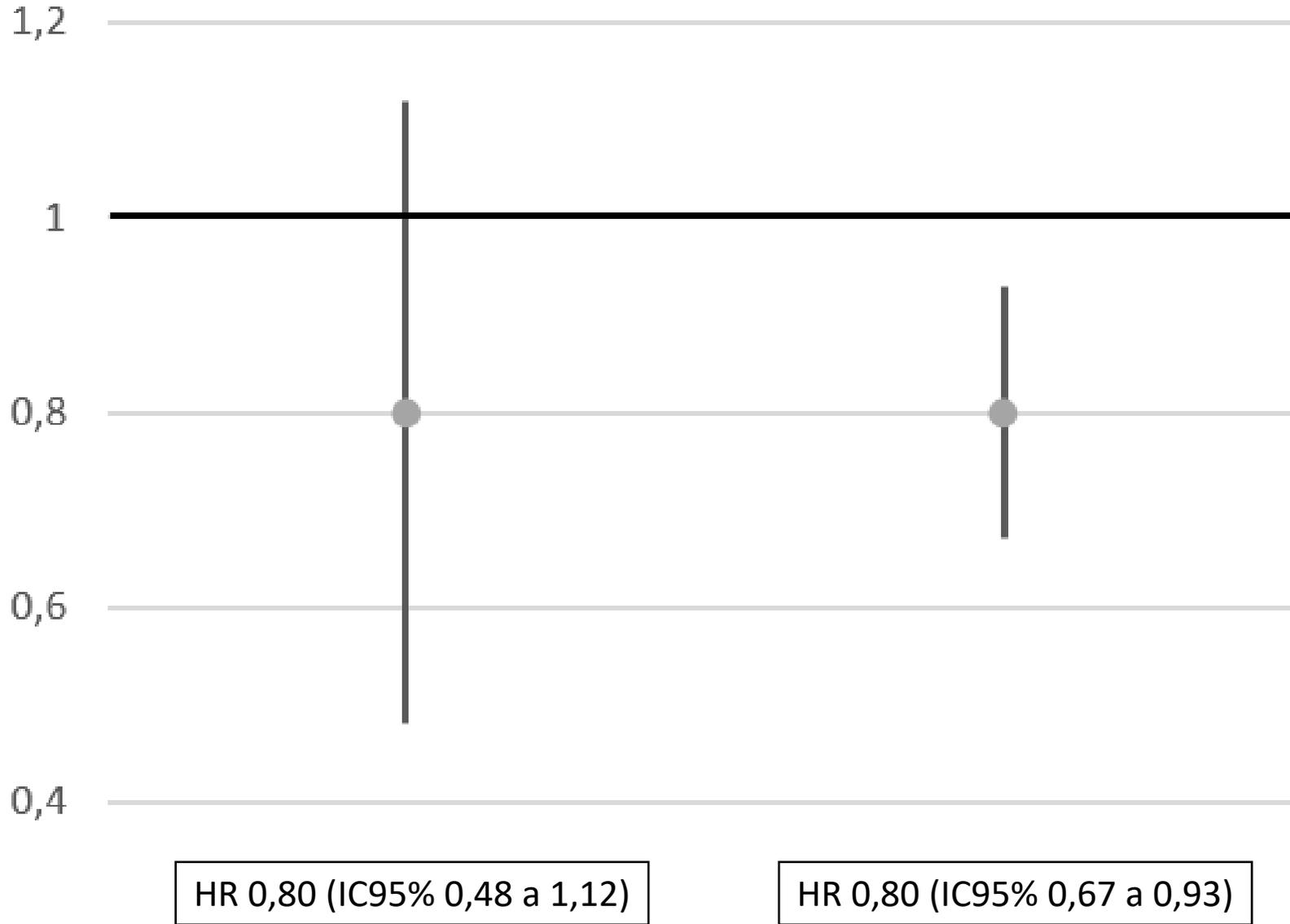
Amplitud del IC es inversamente proporcional a la potencia del estudio y al tamaño muestral

↓ IC ---- ↑ precisión

HR 0,80 (IC95% 0,48 a 1,12)

HR 0,80 (IC95% 0,67 a 0,93)





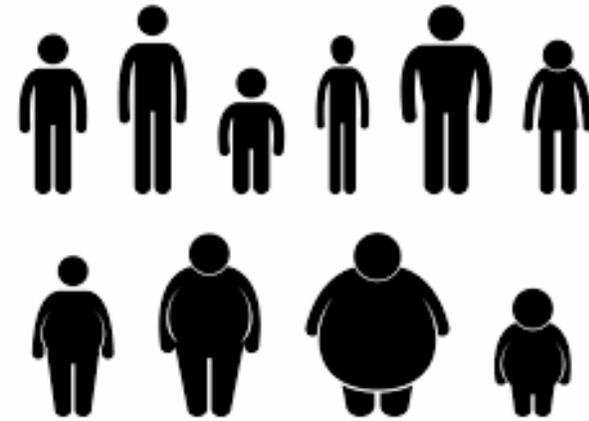
HR 0,80 (IC95% 0,48 a 1,12)

HR 0,80 (IC95% 0,67 a 0,93)

Validez externa



Ensayo clínico



Práctica clínica



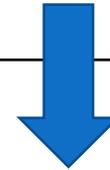
Relación **Validez interna** - **Validez externa**

Eficacia

Probabilidad de que **un promedio de individuos**, se beneficie de un fármaco, bajo condiciones **IDEALES** de actuación establecidas previamente

Efectividad

Probabilidad de que **un individuo**, se beneficie de un fármaco, bajo condiciones **REALES** de aplicación por personal de salud



Características basales pacientes ECA

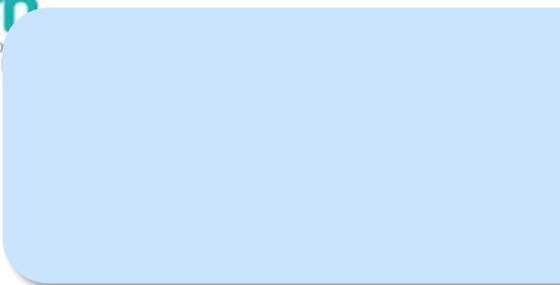
Atención asistencial en nuestro medio

Grado en que los resultados del ECA pueden ser **generalizados** a nuestros pacientes

**PACIENTES
SIMILARES**

Ver criterios de inclusión y exclusión

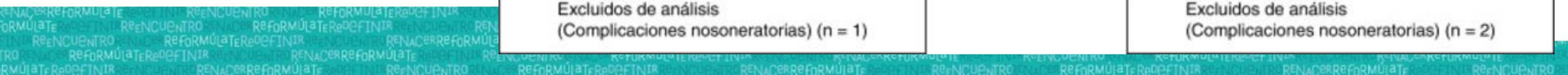
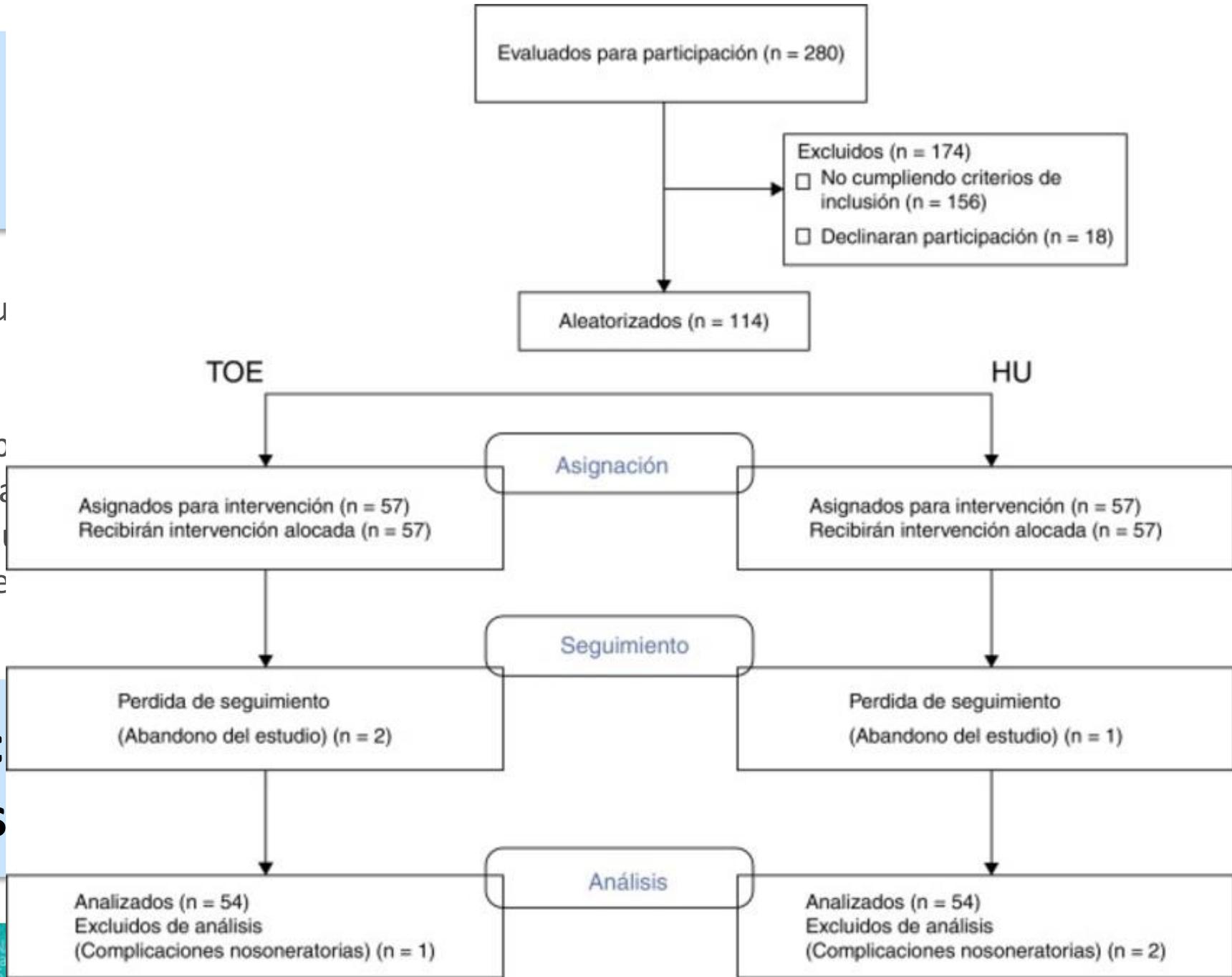


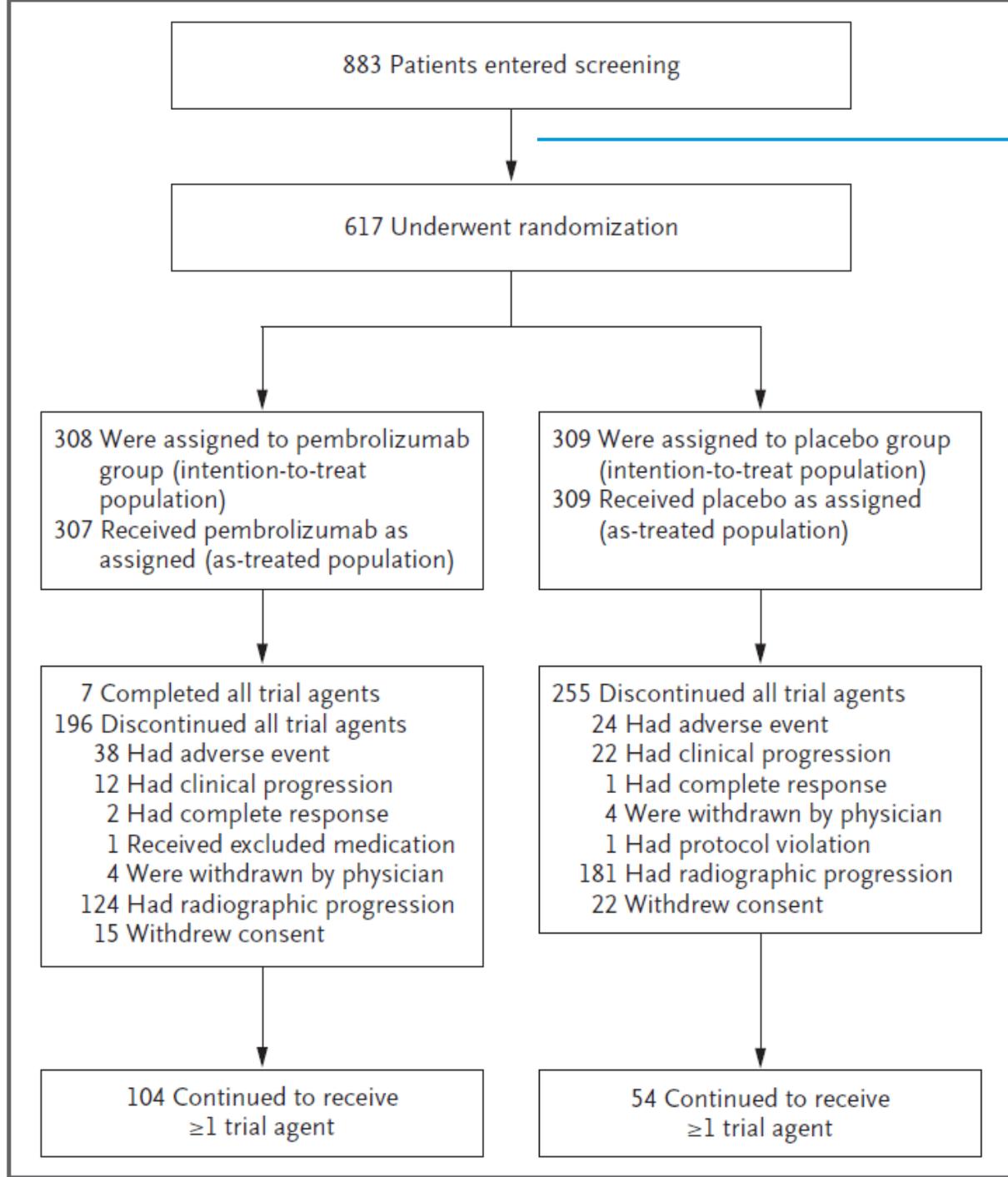


La iniciativa STARD y la guía de diagrama de flujo:

- Nº participantes elegibles
- Nº de pérdidas en cada grupo
- Nº participantes que completaron el estudio
- Nº de pacientes que recibieron la intervención

Proporción pacientes excluidos





Excluidos 266
(30,12%)

??



FLUJO DE PACIENTES

La iniciativa STARD y la guía CONSORT 2010 → altamente recomendable la presentación de diagrama de flujo:

- Nº participantes elegibles inicialmente
- Nº de pérdidas en cada etapa
- Nº participantes que cumplieron los criterios de inclusión
- Nº de pacientes que recibieron o no la intervención a estudio

- TRANSPARENCIA
- POSIBLES SESGOS DE SELECCIÓN

**Proporción pacientes
excluidos**

- ✓ Exclusión de grupo de pacientes, ¿JUSTIFICADO?
- ✓ < 20%
- ✓ CONCLUSIONES DEL ESTUDIO

COMPARADOR

- Definición de la enfermedad
- Criterios de selección de pacientes
- Comparador

¿adecuadas al problema de salud de interés?



Placebo: adecuado?

Activo: estándar?

SEGURIDAD

A tener en cuenta de los EC

= Eficacia

- ❖ Diseño de estudio adecuado (comparativo), aleatorizado, muestra suficiente
- ❖ Análisis estadístico adecuado:
 - Valor de p, potencia del estudio, intervalo de confianza. Difícil para detectar RAM raras o muy raras
- ❖ Fuentes de financiación del estudio y revisión

SEGURIDAD

NCI Common Terminology Criteria for Adverse Events (CTCAE)

Grado 1	Evento adverso leve
Grado 2	Evento adverso moderado
Grado 3	Evento adverso grave
Grado 4	Evento adverso con riesgo de mortalidad o de discapacidad
Grado 5	Muerte asociada con un evento adverso

Distinguir lo importante → Frecuencia
Gravedad

Seguridad asociada o no al efecto terapéutico

Silvia Fénix Caballero

silvia.fenix.sspa@juntadeandalucia.es



Agradecimientos:

Jorge Díaz Navarro

Emilio J. Alegre del Rey

¡Gracias!

