Conceptos estadísticos clave en el diseño

Guillermo Villacampa

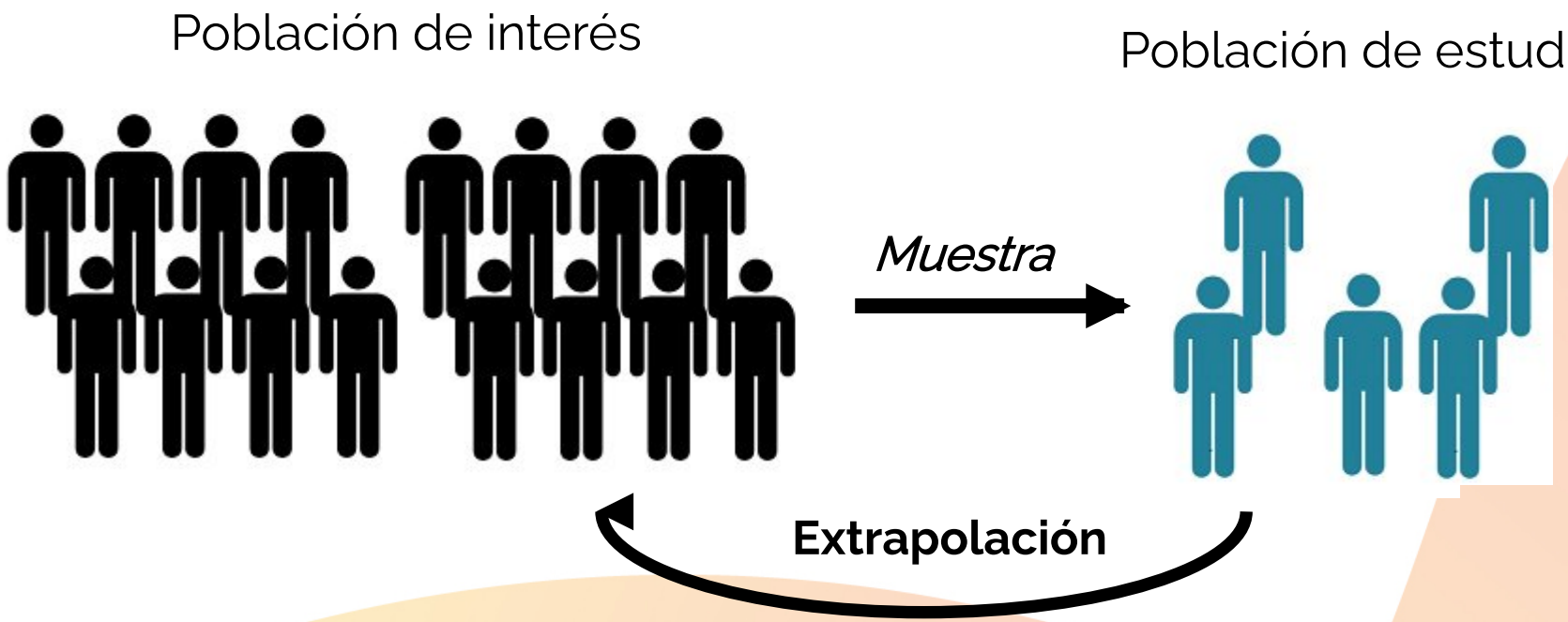
Vall d'Hebron Insititute of Oncology (VHIO), Barcelona

Organizado por: GUARD CONSORTIUM

Resumen

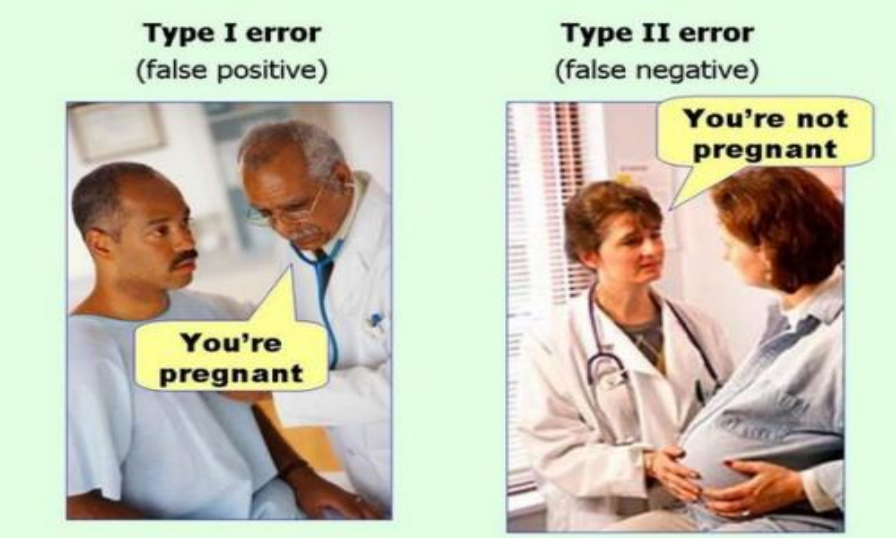
- Plan estadístico y **cálculo del tamaño muestral**
- **Análisis de supervivencia** y endpoints subrogados
- Diseño de estudios **fase III**
- Estudios con **real-world data** y **calidad de los datos**

Conceptos básicos en bioestadística



Conceptos básicos en bioestadística

- Control de errores:
  - *Error de falso positivo* (error tipo I) → *Alpha*
  - *Error de falso negativo* (error tipo II) → *Beta (potencia estadística)*



GUARD CONSORTIUM

Plan de análisis estadístico

1. **Población a analizar**
   - ➢ *Intención de tratar/ Por protocolo/ Análisis de sensibilidad/ Población por biomarcador*
2. **Definición de los endpoints**
   - ➢ Eficacia/seguridad/calidad de vida, etc
3. **Métodos de análisis**
4. **Test de hipótesis**
5. **Cálculo del tamaño muestral**

Tamaño muestral (fase I)

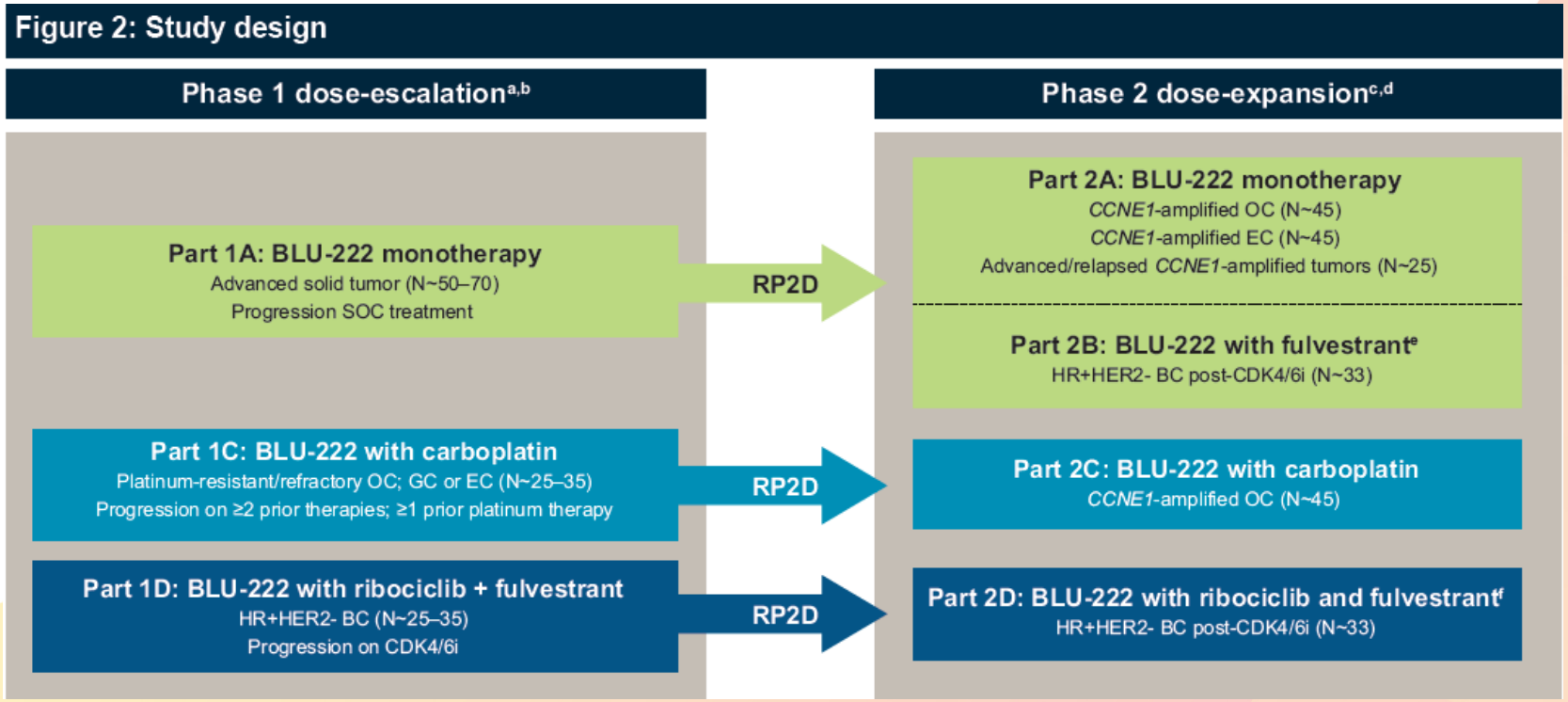
En diseños fase I clásicos (3+3, etc), la **N final** depende de las **toxicidades observadas** y del **número de dosis**. Se puede fijar un **N máxima.**

GUARD CONSORTIUM

Tamaño muestral (expansión de dosis)

491TiP

A first-in-human phase 1/2 study of BLU-222, a potent, selective CDK2 inhibitor in patients with *CCNE1*-amplified or CDK4/6 inhibitor-resistant advanced solid tumors



GUARD CONSORTIUM

Reference: VELA trial, Yap et al, ESMO2022

Tamaño muestral (fase II)

Mucha **variabilidad** dependiendo el diseño del estudio y si existe aleatorización.

GUARD CONSORTIUM

Tamaño muestral (fase II)

Avelumab + sacituzumab govitecan vs avelumab monotherapy as first-line maintenance treatment in patients with advanced urothelial carcinoma: interim analysis from the JAVELIN Bladder Medley phase 2 trial

Jean Hoffman-Censits,[1] Marinos Tsiatas,[2] Peter Mu-Hsin Chang,[3,4] Miso Kim,[5] Lorenzo Antonuzzo,[6,7] Sang Joon Shin,[8] Georgios Gakis,[9] Normand Blais,[10] Se Hyun Kim,[11] Annabel Smith,[12] José Angel Arranz Arija,[13] Yu Li Su,[14] Flora Zagouri,[15] Marco Maruzzo,[16] Christophe Tournigand,[17] Frédéric Forget,[18] Astrid Schneider,[19] Karin Tyroller,[20] Natalia Jacob,[19] Begoña Pérez Valderrama[21]

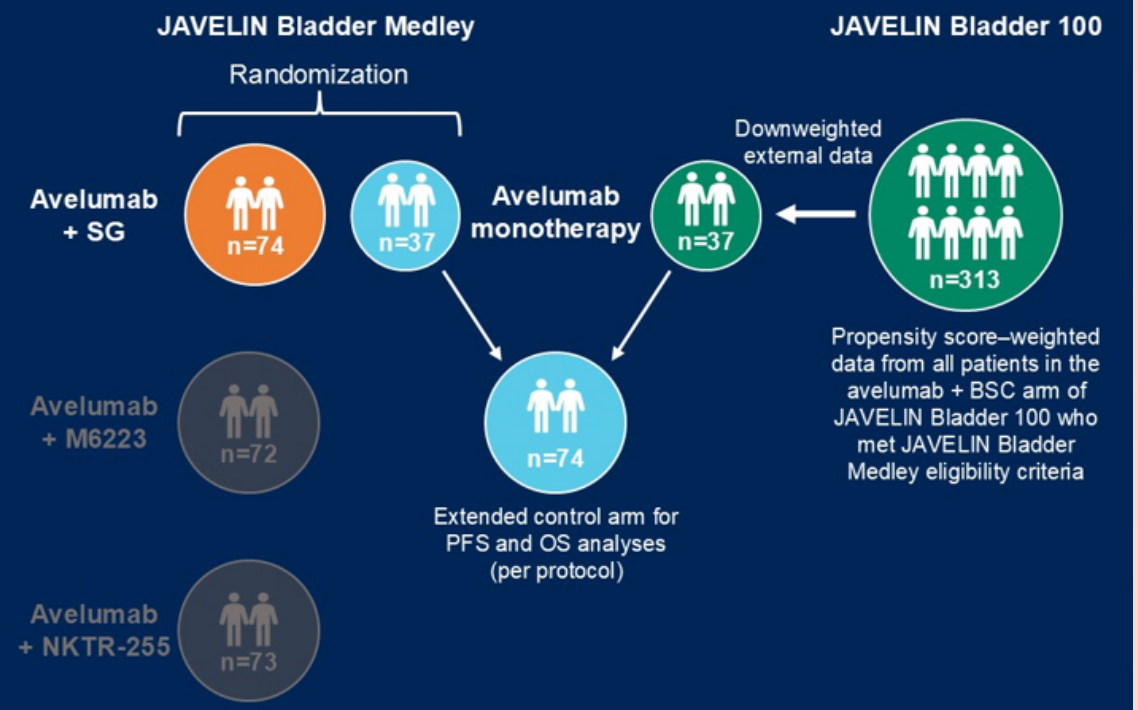
[1]The Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins, Johns Hopkins Medical Institutions, Baltimore, MD, USA; [2]Athens Medical Center, Marousi, Greece; [3]Taipei Veterans General Hospital, Taipei, Taiwan; [4]Institute of Biopharmaceutical Science, National Yang Ming Chiao Tung University, Hsinchu, Taiwan; [5]Seoul National University Hospital, Seoul National University College of Medicine, Seoul, South Korea; [6]Careggi University Hospital, Florence, Italy; [7]University of Florence, Florence, Italy; [8]Yonsei Cancer Center, Yonsei University College of Medicine, Seoul, South Korea; [9]Martin-Luther-University of Halle-Wittenberg, Halle, Germany; [10]University of Montreal Health Centre (CHUM), Montréal, QC, Canada; [11]Seoul National University Bundang Hospital, Seoul National University College of Medicine, Seongnam, South Korea; [12]Icon Cancer Centre, Adelaide, Australia; [13]Hospital Universitario Gregorio Marañón, Madrid, Spain; [14]Kaohsiung Chang Gung Memorial Hospital, Kaohsiung, Taiwan; [15]"Alexandra" General Hospital of Athens, Athens, Greece; [16]IOV—Istituto Oncologico Veneto IRCCS, Padova, Italy; [17]AP-HP, Hôpital Henri Mondor, Créteil, France; [18]Centre Hospitalier de l'Ardenne, Libramont-Chevigny, Belgium; [19]the healthcare business of Merck KGaA, Darmstadt, Germany; [20]EMD Serono, Billerica, MA, USA; [21]Hospital Universitario Virgen del Rocio, Seville, Spain.

GUARD CONSORTIUM

Tamaño muestral (fase II)

Statistical design of the extended control arm

- Per protocol, PFS and OS data in the control arm were extended using external data from the JAVELIN Bladder 100 phase 3 trial[1]
  - Patients who met JAVELIN Bladder Medley eligibility criteria were included (n=313/350)
  - To account for population differences, external patients were propensity-score weighted using predefined prognostic factors*
  - The sum of external patients was downweighted to 37 to be equal to the number of randomized patients



GUARD CONSORTIUM

Tamaño muestral (fase III)

1. ¿Cuál es el **endpoint primario**?
2. ¿En qué **población** se evaluará?
3. **Estimación del beneficio del tratamiento**. ¿Qué tan bueno es el tratamiento experimental en comparación con el control en términos del endpoint primario?
4. ¿Cuál es la mejor **estrategia** para realizar el análisis? Un único análisis final, análisis intermedios, etc.
5. **Tiempo de reclutamiento**.
6. **Perdida de seguimiento** (dropout rate)
7. **Control de errores** (falso positivo y falso negativo)

Estimación del beneficio del tratamiento

Estimación del beneficio del tratamiento:

"Cuanto mayor es el beneficio esperado, menor es el tamaño de la muestra. Cuanto menor es el beneficio esperado, mayor es el tamaño de la muestra."

Beneficio	*Riesgos*	*"fortalezas"*
➢ **Sobreestimación**	Alto riesgo de tener un resultado negativo	Se necesita una N pequeña
➢ **Infraestimación**	Alto riesgo de diseñar un studio no factible	Altas probabilidades de tener un resultado positivo

GUARD CONSORTIUM

Tamaño muestral (fase III)

El tamaño muestral depende de:

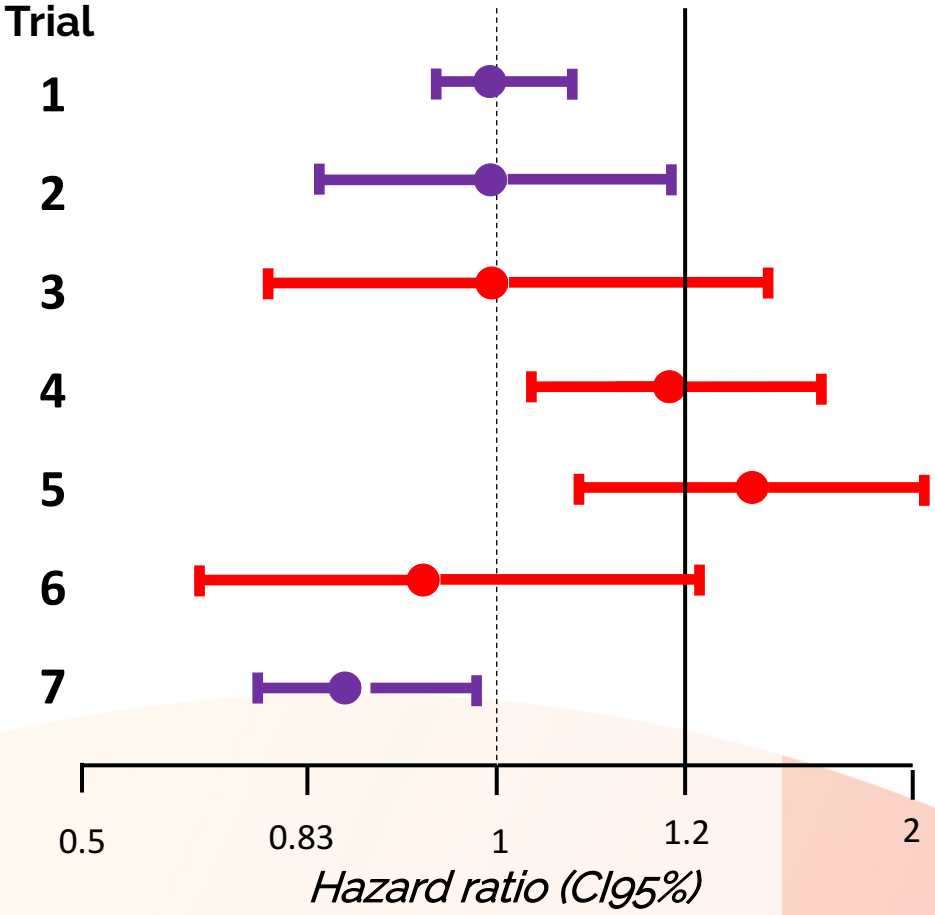
- Error alpha — 5% → Muy controlada
- Potencia estadística — 80%/90%
- Hazard ratio esperada (endpoint primario) — ? → Se realista!
- Ratio de aleatorización — 1:1 / 2:1 → *?*
- Periodo de reclutamiento/duración del estudio — ?
- Ratio de drop-out — ?
- Número de análisis — ?

La importancia de los **eventos**

GUARD CONSORTIUM

Estudios de no-inferioridad

Fármaco	Trial	
A	1	Non-inferiority
A	2	Non-inferiority
A	3	Inferiority
B	4	Inferiority
C	5	Inferiority
D	6	Inferiority
E	7	Non-inferiority



GUARD CONSORTIUM

Estudios de no-inferioridad

Efficacy

At final analysis (PD-L1 CPS of 1 or greater), median OS with pembrolizumab (10.6 months; 95% CI, 7.7-13.8) vs chemotherapy (11.1 months; 95% CI, 9.2-12.8) met the criteria for non-inferiority (HR, 0.91; 99.2% CI, 0.69-1.18; noninferiority margin, 1.2) (**Figure 2**A). The 12-month and 24-month OS rates were

Shitara at al, JAMA Oncol
2021 (K

placebo. Non-inferiority would be claimed if the upper two-sided 95% CI of the breast-cancer recurrence HR was less than the non-inferiority margin of delta=1·278

Kenemans at al, Lancet Oncology
2009 (NCT00408863)

Overall, the non-inferiority margin is fixed between hazard ratio of 1.2 and 1.3

PERSEPHONE showed non-inferiority for 6 months of trastuzumab compared with 12 months. Our definition of non-inferiority was no worse than 3% absolute below the standard group's 4-year disease-free survival, and the non-inferiority limit was thus calculated as an HR of less than 1·32. Notably, the upper confidence limit of the HR

Earl at al, Lancet Oncology
2019 (PERSEPHONE)

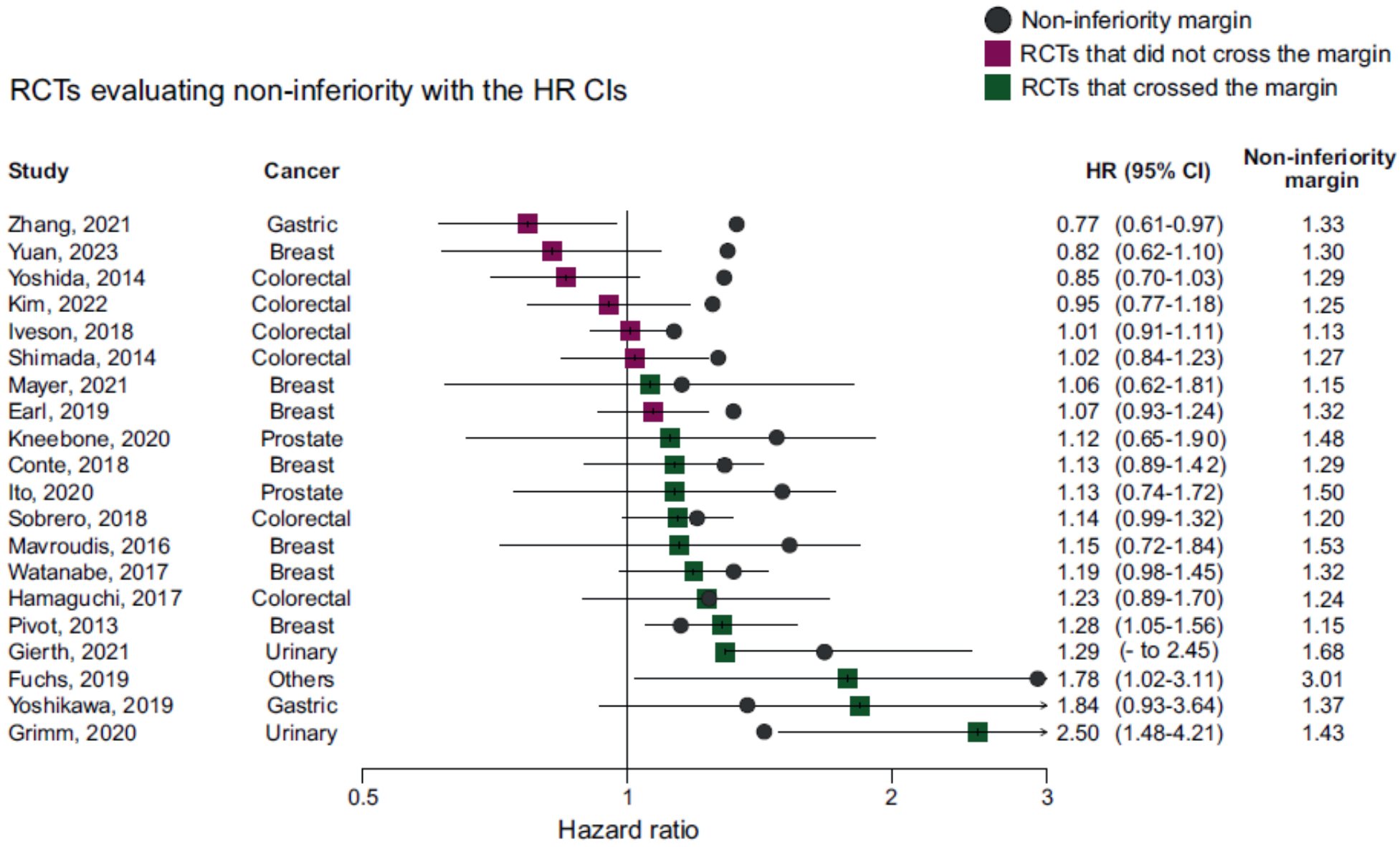
Statistical analysis

Based on the results of previous studies for paclitaxel in breast cancer, the expected median PFS was 5.5 months for PTX and 6.35 months for NK105. Then, assuming a randomisation period of 18 months, a follow-up period of 12 months, a one-sided significance level of 2.5%, a power of 85% and the above-described non-inferiority margin of 1.215, it was estimated that

Fujiwara at al, BJC 2019
(NCT01644890)

GUARD CONSORTIUM

Tamaño muestral: No-inferioridad



The ty

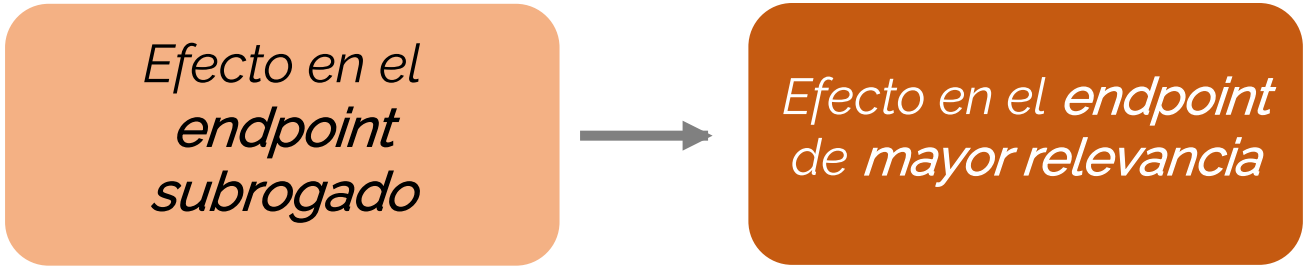
Ian F Tannock, M

Opportunitie
shorter dura
approved reg
alternative ar
superiority. N
approved the
depending o
such as toxic
be required t
longer durat
inferiority tr
intensive the
large burden
superiority o
simply as "cc

Resumen

- Plan estadístico y **cálculo del tamaño muestral**
- **Análisis de supervivencia** y endpoints subrogados
- Diseño de estudios **fase III**
- Estudios con **real-world data** y **calidad de los datos**

Endpoints subrogados



- ***Ejemplos de endpoint subrogados***: *Event-free survival (EFS), pathological complete response (pCR), y objective response rate (ORR)*

Endpoints subrogados

Asociación débil entre los **endpoints subrogados** y la **supervivencia global** en el contexto de la **inmunoterapia** y las terapias dirigidas

Irreconcilable Differences: The Divorce Between Response Rates, Progression-Free Survival, and Overall Survival

Margret Merino, MD[1]; Yvette Kasamon, MD[1]; Marc Theoret, MD[1,2]; Richard Pazdur, MD[1,2]; Paul Kluetz, MD[1,2]; and Nicole Gormley, MD[1]

Merino et al, JCO. 2023

Exposure to US Cancer Drugs With Lack of Confirmed Benefit After US Food and Drug Administration Accelerated Approval

Between 2009 and 2022, the FDA approved 48 under the Accelerated Approval (AA) program based on surrogate endpoints. Fifteen indications (23%) have been withdrawn due to lack of benefit over standard of care.

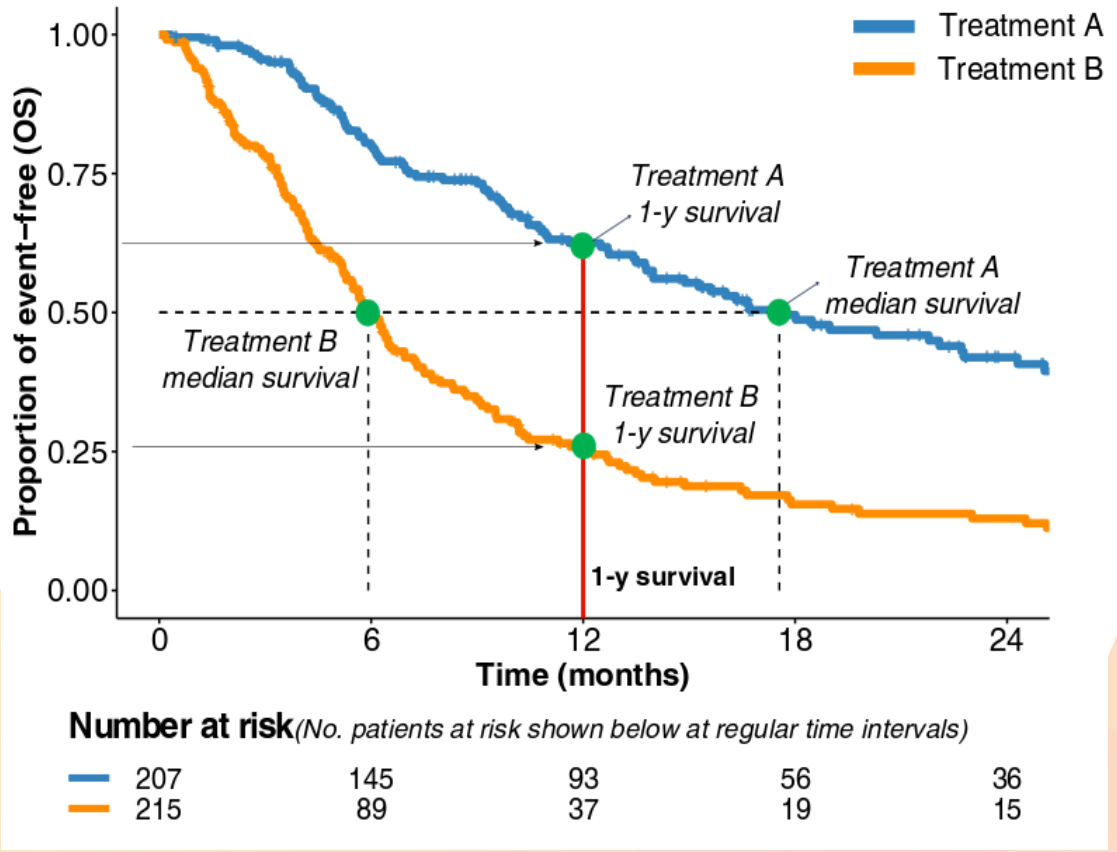
Parikh et al, JAMA Onc. 2023

GUARD CONSORTIUM

Análisis de supervivencia

Tres conceptos claves:

1. **Curvas Kaplan-Meier**
2. **Hazard ratio (HR)**
3. **Test de log-rank**



HR=0.35 *(CI95% 0.20 – 0.55), p<0.001*

Risk reduction: 1 – 0.35 = 65%

GUARD CONSORTIUM

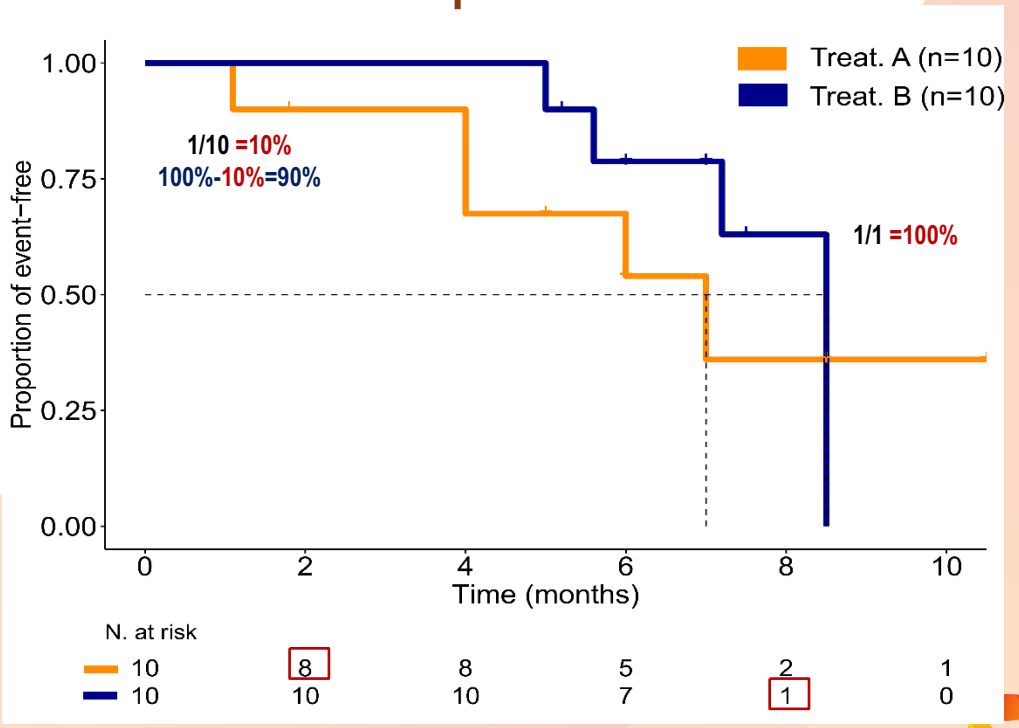
Curvas Kaplan-Meier

Ejemplo: N= 20 pacientes (10 con tratamiento A y 10 con tratamiento B)

Base de datos

ID	Treatment	time	status
1	A	1.1	1
2	A	1.8	0
3	A	4.0	1
4	A	4.0	1
5	A	5.0	0

Kaplan-Meier



GUARD CONSORTIUM

Efecto del tratamiento (más allá del p-valor)

El intervalo de confianza es más informativo que el p-valor

Study	HR (CI95%)	P-value
EV-302 trial (urothelial carcinoma)	**0.45 (0.38 – 0.54)**	
ATLAS (kidney cancer)	**0.87 (0.66 – 1.45)**	
APHINITY (breast cancer)	**0.81 (0.66 – 1.00)**	

GUARD CONSORTIUM

Efecto del tratamiento (más allá del p-valor)

El intervalo de confianza es más informativo que el p-valor

Study	HR (CI95%)	P-value
EV-302 trial (urothelial carcinoma)	**0.45 (0.38 – 0.54)**	**<0.001**
ATLAS (kidney cancer)	**0.87 (0.66 – 1.45)**	**0.32**
APHINITY (breast cancer)	**0.81 (0.66 – 1.00)**	**0.05**

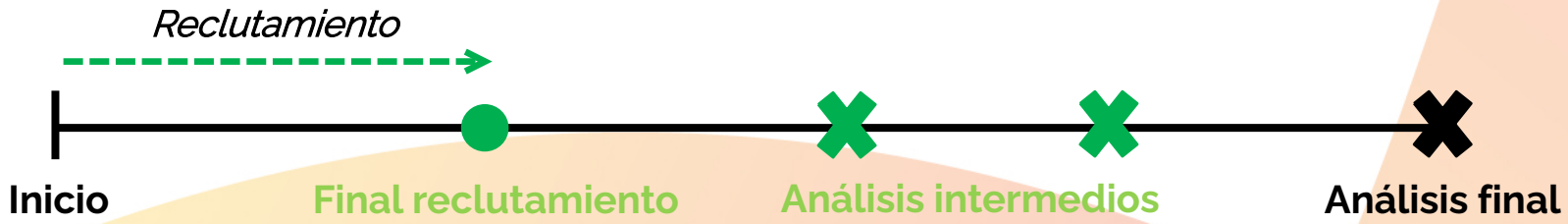
Resumen

- Plan estadístico y **cálculo del tamaño muestral**
- **Análisis de supervivencia** y endpoints subrogados
- Diseño de estudios **fase III**
- Estudios con **real-world data** y **calidad de los datos**

Fase III: Diseños secuenciales

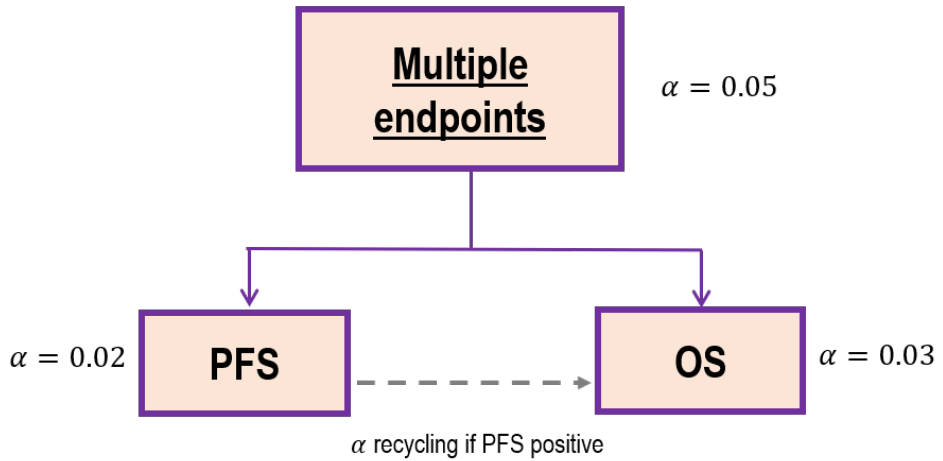
1. Diseño clásico

2. Diseño secuencial

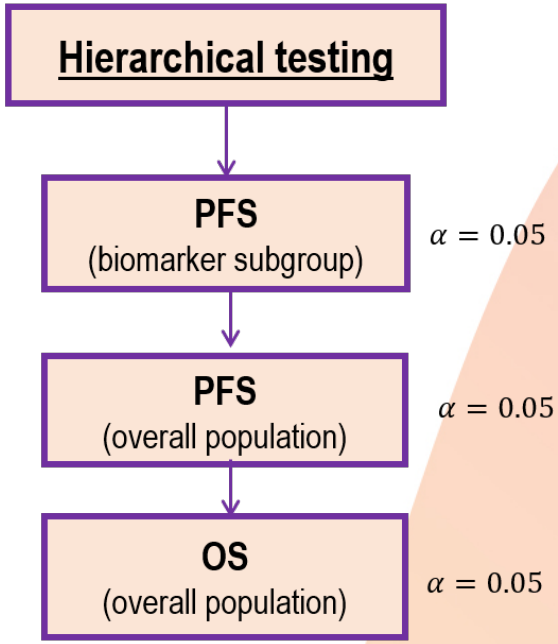


GUARD CONSORTIUM

Endpoints en fase III



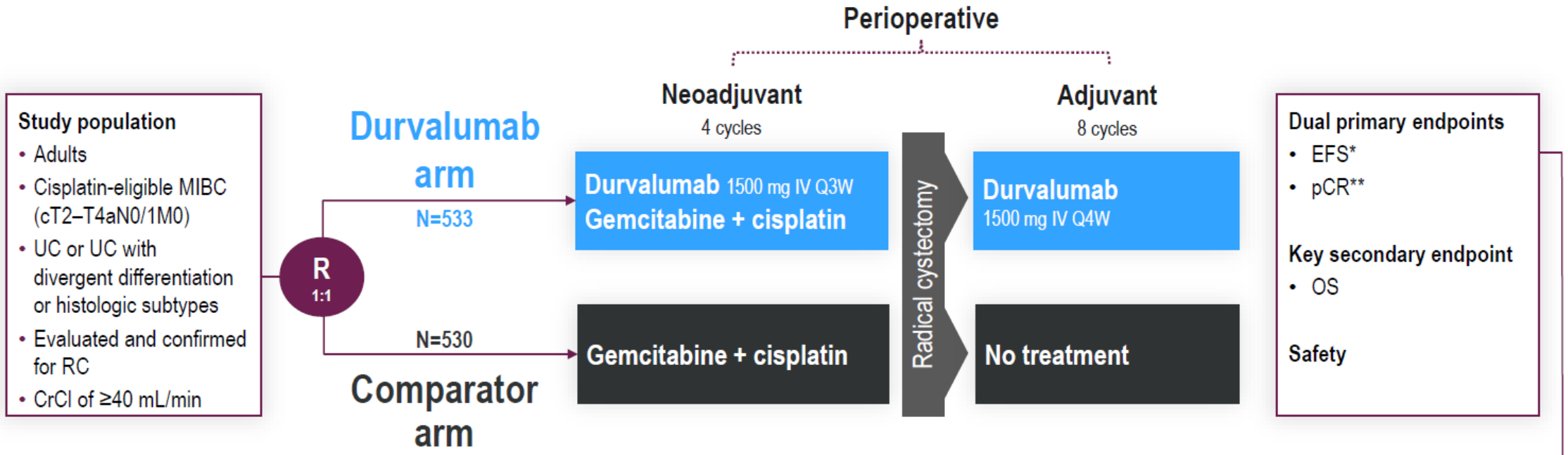
- Alpha splitting between endpoints.
- If one endpoint is positive the study is positive.
- If the first tested endpoint is positive, alpha recycling can be performed



- Only if the previous endpoint is positive the next endpoint can be tested.

GUARD CONSORTIUM

Estudio NIAGARA



Stratification factors

Clinical tumour stage (T2N0 vs >T2N0)

Renal function (CrCl ≥60 mL/min vs ≥40–<60 mL/min)

PD-L1 status (high vs low/negative expression)

Gemcitabine/cisplatin dosing

CrCl ≥60 mL/min: Cisplatin 70 mg/m^2 + gemcitabine 1000 mg/m^2 Day 1, then gemcitabine 1000 mg/m^2 Day 8, Q3W for 4 cycles

CrCl ≥40–<60 mL/min: Split-dose cisplatin 35 mg/m^2 + gemcitabine 1000 mg/m^2 Days 1 and 8, Q3W for 4 cycles

EFS was defined as:

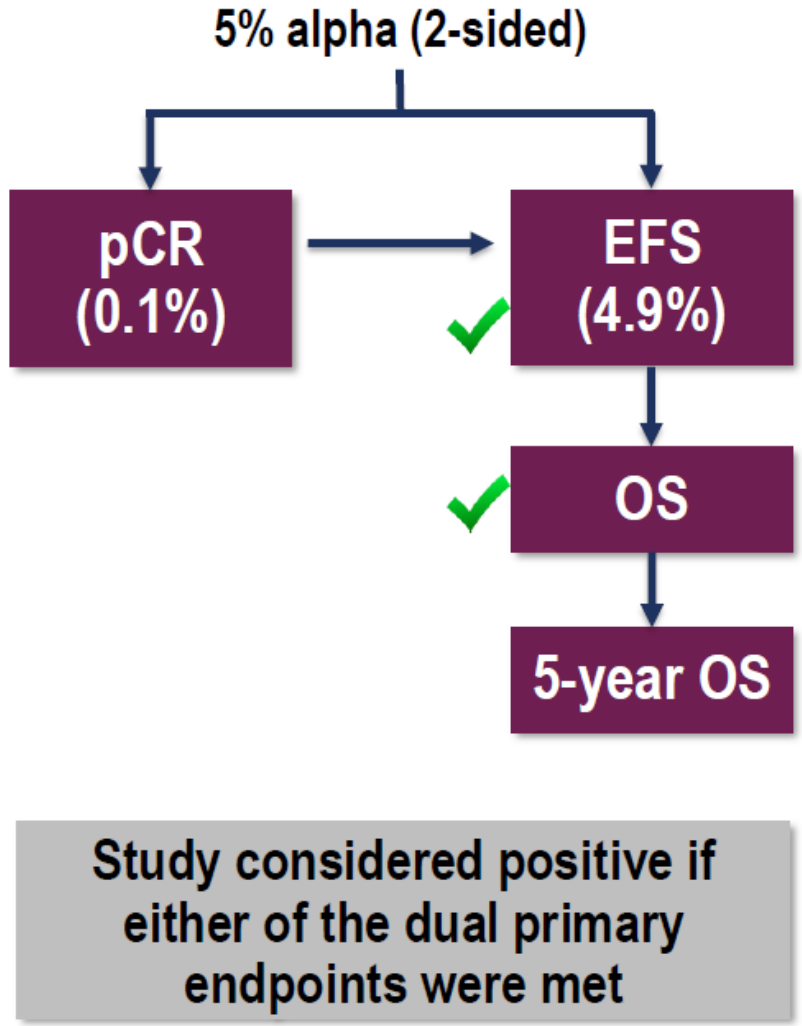
- Progressive disease that precluded RC
- Recurrence after RC
- Date of expected surgery in patients who did not undergo RC
- Death from any cause

Other endpoints (not reported here): DFS, DSS, MFS, HRQoL, 5-year OS

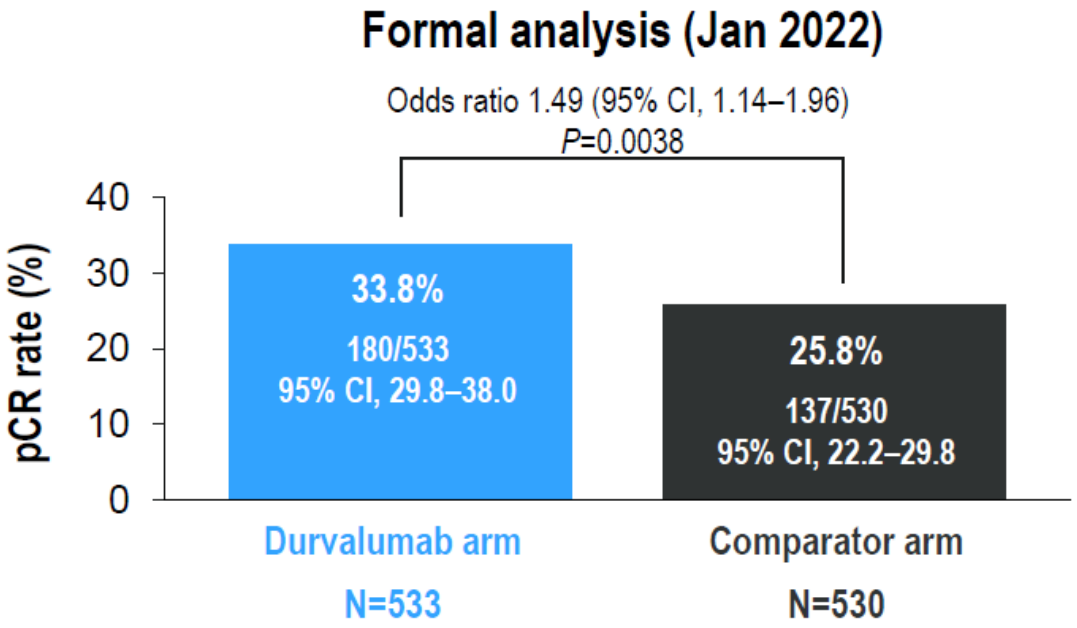
Estudio NIAGARA

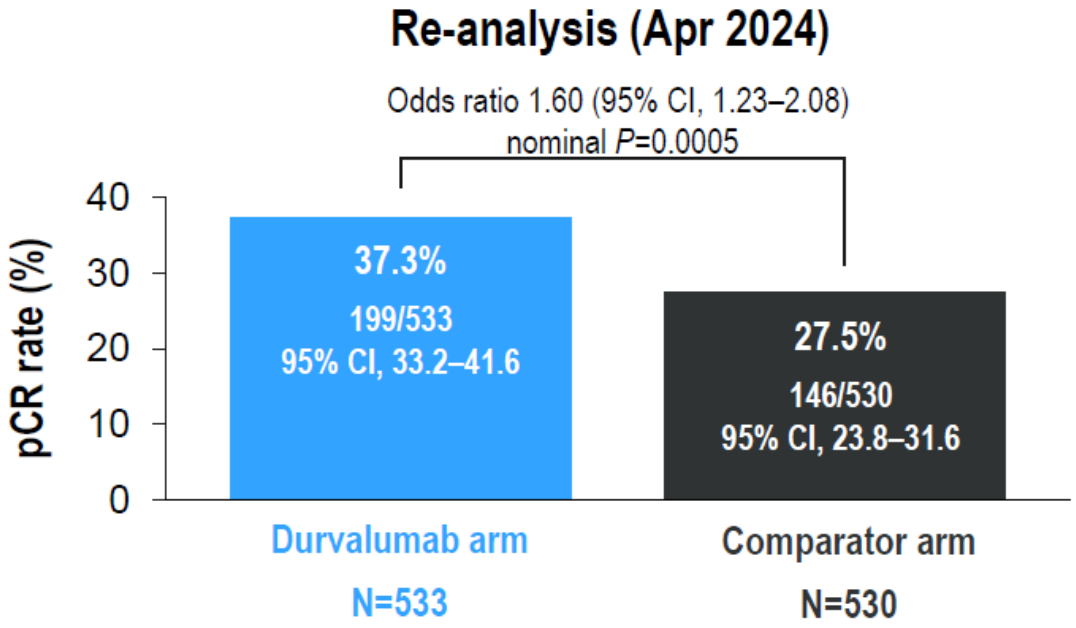
- A multiple testing procedure with an **alpha-exhaustive recycling** strategy and gatekeeping strategy was used across the **dual primary endpoints** and then the secondary endpoints of OS and 5-year OS

- **EFS analysis**: 451 EFS events were needed to provide the trial with 90% power to detect a significant between-group difference in event-free survival, with an underlying hazard ratio of 0.73 and a two-sided alpha level of 0.049.



pCR endpoint





- The planned formal analysis for pCR was not statistically significant (threshold for significance, p-value 0.001)
- 59 evaluable samples were incorrectly considered non-responders rather than their true result*

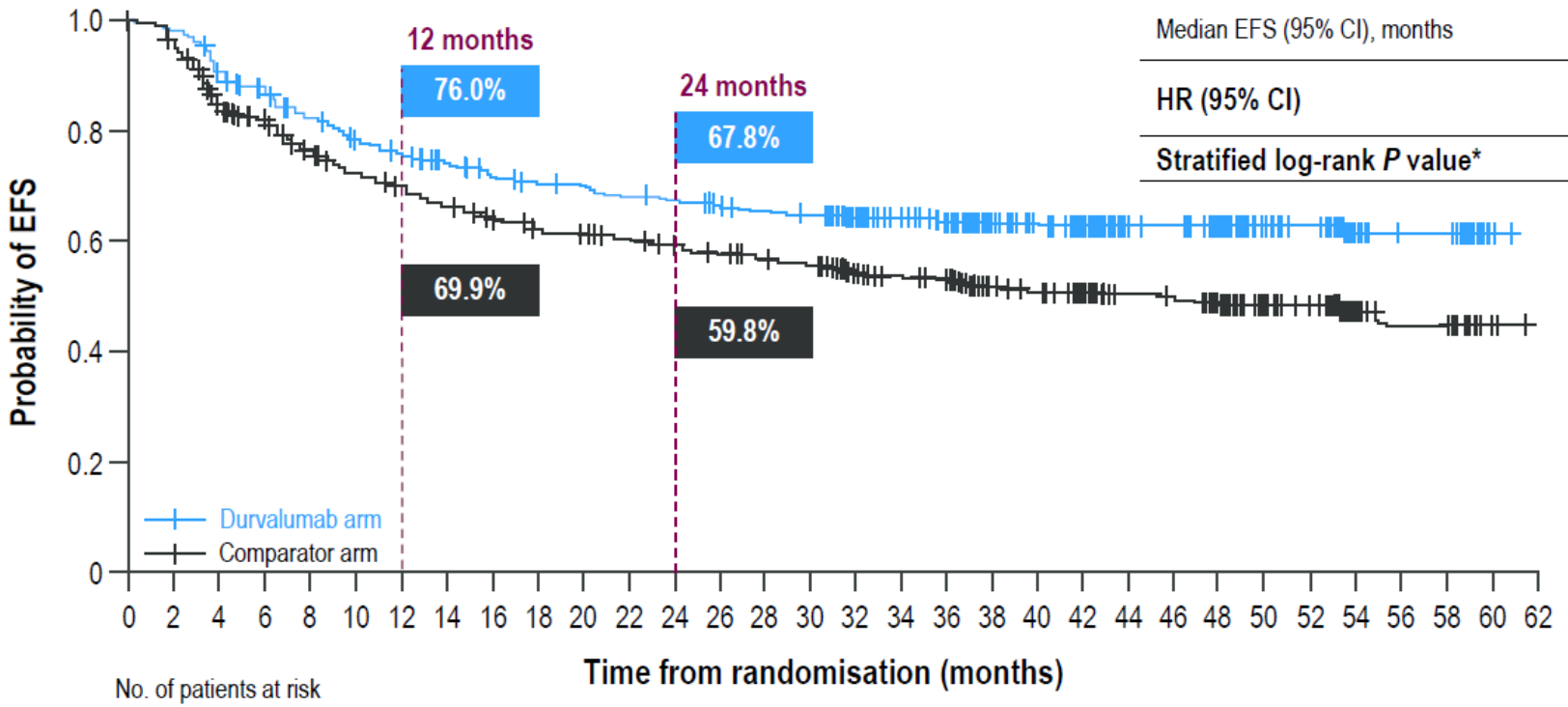
- The re-analysis showed nominal statistical significance in favour of the durvalumab arm
- This analysis includes the results of the 59 omitted samples (28 additional pCRs)*

EFS endpoint

NIAGARA: Event-free Survival by Blinded Independent Central Review (ITT)

BARCELONA 2024 ESMO congress

	Durvalumab arm N=533	Comparator arm N=530
Number of events, n (%)	187 (35.1)	246 (46.4)
Median EFS (95% CI), months	NR (NR–NR)	46.1 (32.2–NR)
HR (95% CI)	0.68 (0.56–0.82)	
Stratified log-rank *P* value*	<0.0001	



Median follow-up in censored patients: 42.3 months (range, 0.03–61.3)

No. of patients at risk

Durvalumab arm	533	519	475	454	424	401	386	370	356	348	344	335	330	321	315	312	282	269	255	214	202	180	141	140	115	86	81	32	20	20	1	0
Comparator arm	530	498	437	416	381	358	343	328	313	300	296	288	281	273	264	259	228	219	214	177	172	159	132	129	94	69	62	24	18	16	2	0

Ensayos pragmáticos

The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812 SEPTEMBER 18, 2025 VOL. 393 NO. 11

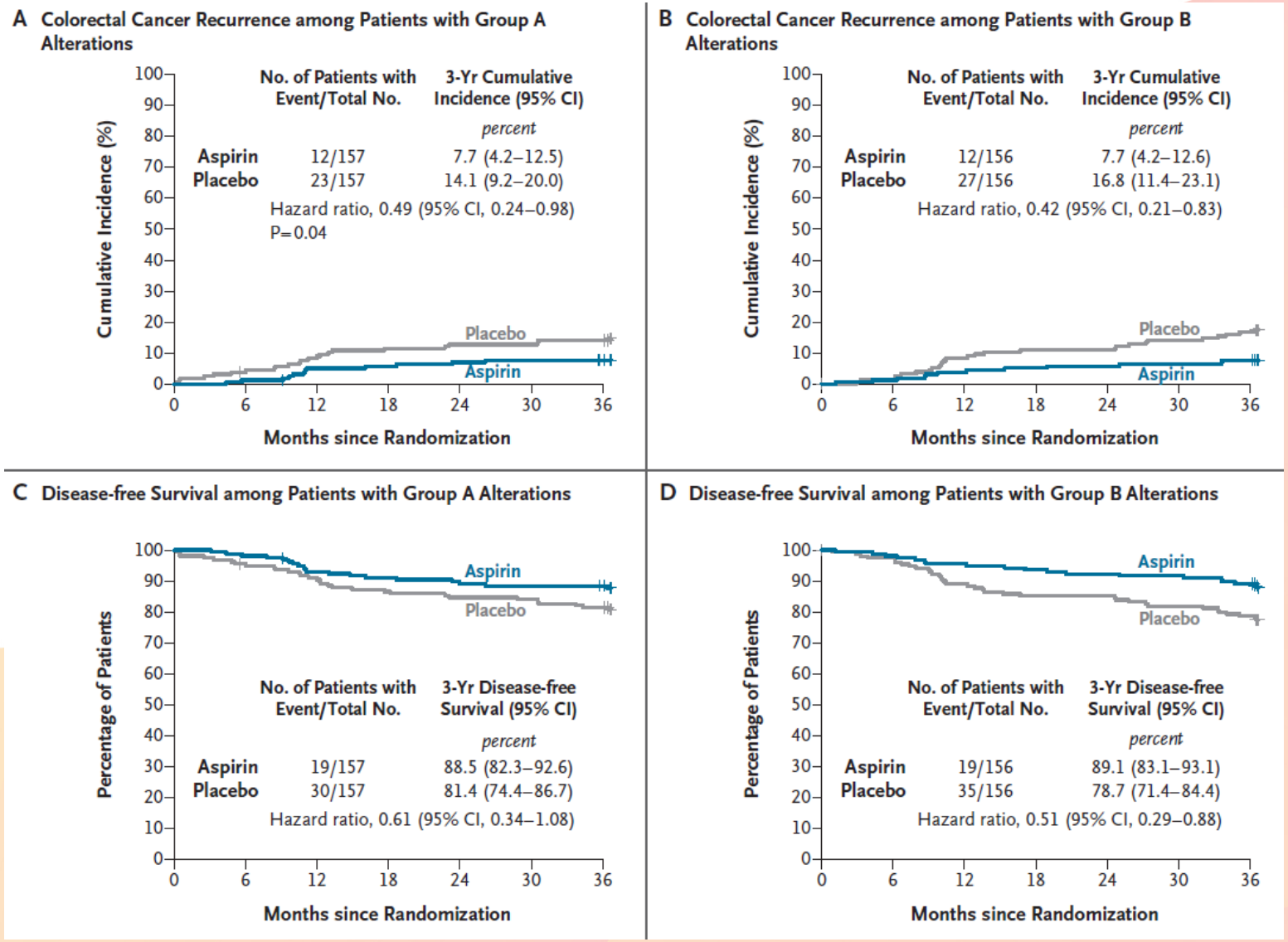
Low-Dose Aspirin for PI3K-Altered Localized Colorectal Cancer

A. Martling,[1,2] I. Hed Myrberg,[3] M. Nilbert,[4] H. Grönberg,[5] F. Granath,[3] M. Eklund,[5] T. Öresland,[6,7] L.H. Iversen,[8] C. Haapamäki,[9] M. Janson,[10] K. Westberg,[1,11] J. Segelman,[1,12] U. Ersson,[13] M. Prytz,[14,15] E. Angenete,[15,16] R. Bergström,[5] M. Mayrhofer,[5,17] B. Glimelius,[18] and J. Lindberg,[19] for the ALASCCA Study Group*

GUARD CONSORTIUM

Ensayos pragmáticos

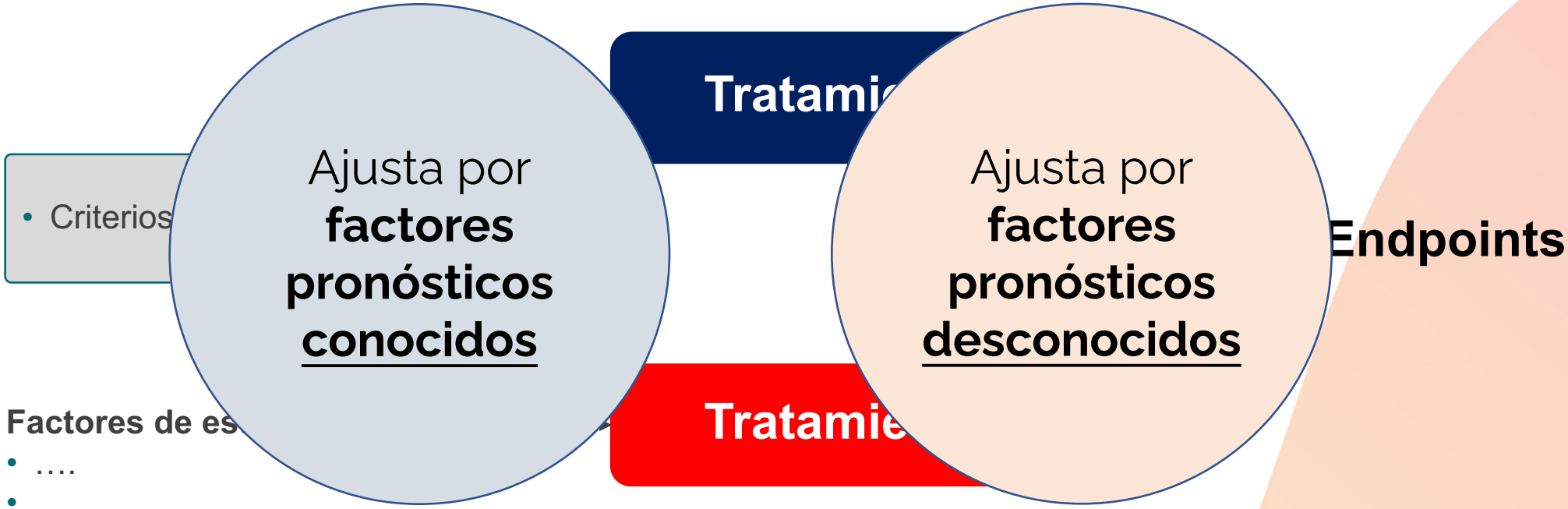
Patients were recruited from April 6, 2016, through July 19, 2021, with follow-up evaluations conducted every 3 months in person or by telephone. Surveillance for colorectal cancer recurrence included thoracic and abdominal computed tomography or magnetic resonance imaging in accordance with national guidelines, which recommended imaging 1 year and 3 years after surgery. Symptomatic or suspected cases of recurrence were evaluated at the local investigator's discretion. Adherence to the trial regimen was assessed by means of pill counts at clinic visits and patient-reported adherence during follow-up telephone calls.



Resumen

- Plan estadístico y **cálculo del tamaño muestral**
- **Análisis de supervivencia** y endpoints subrogados
- Diseño de estudios **fase III**
- Estudios con **real-world data** y **calidad de los datos**

Ensayos clínicos aleatorizados



GUARD CONSORTIUM

Limitaciones de los ensayos clínicos

1. **Sesgo de selección** en la población incluida
2. Requier
3. **Costos**
4. **Seguimiento** largo
5. Falta de adherencia
6. Etc.

Estudios de vida real (RWD)

GUARD CONSORTIUM

¿Para qué podemos utilizar los estudios de vida real?

1. **Confirmar la eficacia** observado en los ensayos clínicos.
2. Evaluar el efecto de un fármaco en **subgrupos** de **pacientes no incluidos** en el estudio pivotal.
3. Comparaciones **head-to-head** entre tratamientos aprobados.
4. Definir la efectividad **cuando** no existen **RCTs** “tradicionales”, o **no son éticos**.
5. **Evaluar** de forma **secuencial** el efecto de **cambios en la práctica clínica** (aprendizaje rápido) – los ensayos clínicos son demasiado lentos.
6. Definir mejor la **seguridad**, incluyendo **efectos adversos tardíos.**

GUARD CONSORTIUM

Limitaciones de estudios de vida real

- No permite ajustar por **factores pronósticos desconocidos.**
- **Datos faltantes** (missing data).
- **Calidad** de los datos.
- **Acceso** limitado a los datos en ciertas situaciones.
- Potenciales errores en los **análisis estadísticos.**
- Etc.

Calidad de los datos

Raising the Bar for Real-World Data in Oncology: Approaches to Quality Across Multiple Dimensions

Emily H. Castellanos, MD, MPH[1]; Brett K. Wittmershaus, BSE[1]; and Sheenu Chandwani, MPH, PhD[1]

DOI https://doi.org/10.1200/CCI.23.00046

ABSTRACT

ACCOMPANYING CONTENT

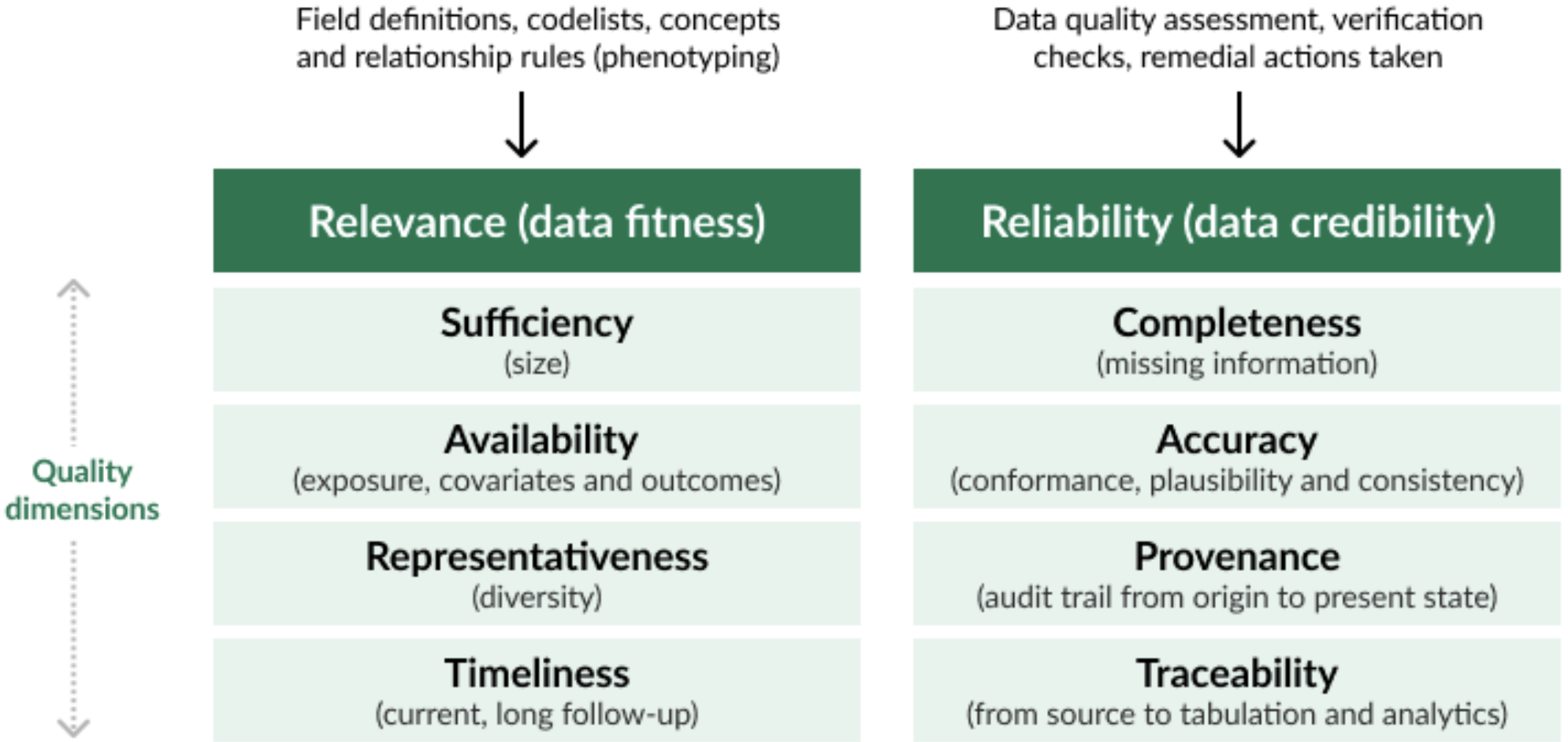
Real-World Database Studies in Oncology: A Call for Standards

Scott D. Ramsey, MD, PhD[1]; Arzu Onar-Thomas, PhD[2]; and Stephanie B. Wheeler, PhD, MPH[3]

DOI https://doi.org/10.1200/JCO.23.02399

GUARD CONSORTIUM

Calidad de los datos

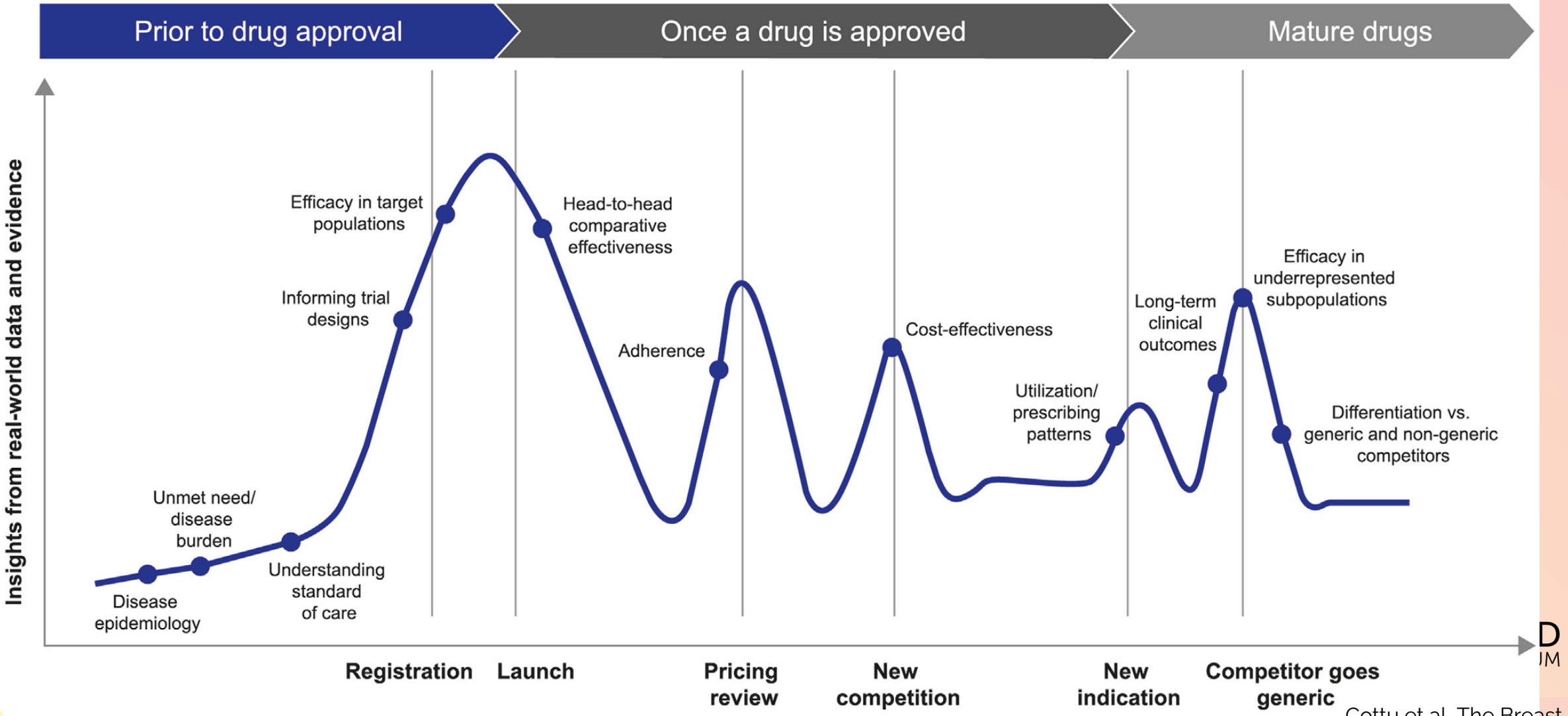


¿Son estos los datos que necesito?

¿Son los datos correctos?

GUARD CONSORTIUM

¿Cuándo de utiliza el RWD?



Cottu et al, The Breast 2022

Guía ESMO estudios RWD (GROW)

ESMO GOOD SCIENCE BETTER MEDICINE BEST PRACTICE

ESMO REAL WORLD DATA AND DIGITAL ONCOLOGY

SPECIAL ARTICLE

ESMO Guidance for Reporting Oncology real-World evidence (GROW)

L. Castelo-Branco[1*], A. Pellat[2,3†], D. Martins-Branco[1,4†], A. Valachis[5†], J. W. G. Derksen[6†], K. P. M. Suijkerbuijk[7], U. Dafni[8,9], T. Dellaporta[9], A. Vogel[10,11,12], A. Prelaj[13,14], R. H. H. Groenwold[15], H. Martins[16], R. Stahel[17], J. Bliss[18], J. Kather[19,20], N. Ribelles[21], F. Perrone[22], P. S. Hall[23], R. Dienstmann[24,25], C. M. Booth[26,27], G. Pentheroudakis[1‡], S. Delaloge[28‡] & M. Koopman[7‡]

GUARD CONSORTIUM

Guía ESMO estudios RWD (GROW)

Methods
3.7: Provide details and timings of **source** and study **data management**. Consider specifying methods of raw data collection, updates and completeness, **data extraction**, cleaning and/or quality **controls** and validation.

Methods
3.15: Specify the **pre-planned strategies** to identify and **mitigate** the main sources of **bias.**

Results
4.1: **Provide the number of cases excluded or nonparticipating and reasons** at each stage of sample selection, as well as numbers lost to follow-up. Compare the cases excluded with those included in the analyses .

GUARD CONSORTIUM

Endpoints en estudios de RWD

Endpoint and definition	Strengths	Limitations
rw-Time to Next Treatment (rwTTNT) Length of time from the index treatment date to the date the patient received their next systemic therapy (next-line therapy) or date of death. If patient has not received subsequent treatment or has not died, patients will be censored at their last known activity or end of follow-up.	• A proxy for disease control or potential benefit • Integrates total amount of time a patient is treated on therapy. • Captures possible response durability of initial treatment. • Advantage over rwTTD given the data availability of timing of next therapy initiation compared to discontinuation. • Not subject to bias associated with variable tumor burden assessments	• Endpoint as defined is specific to next line systemic therapy and is not inclusive of other interventions such as surgery or radiation which could result in bias. • Missingness of data if patients receive treatment outside of the system

Ejemplos de RWE: Trial emulation

JAMA | Original Investigation

Emulation of Randomized Clinical Trials With Nonrandomized Database Analyses
Results of 32 Clinical Trials

Shirley V. Wang, PhD, ScM; Sebastian Schneeweiss, MD, ScD; and the RCT-DUPLICATE Initiative

IMPORTANCE Nonrandomized studies using insurance claims databases can be analyzed to produce real-world evidence on the effectiveness of medical products. Given the lack of baseline randomization and measurement issues, concerns exist about whether such studies produce unbiased treatment effect estimates.

OBJECTIVE To emulate the design of 30 completed and 2 ongoing randomized clinical trials (RCTs) of medications with database studies using observational analogues of the RCT design parameters (population, intervention, comparator, outcome, time [PICOT]) and to quantify agreement in RCT-database study pairs.

Funding/Support: This study was funded by contracts HHSF223201710186C and HHSF223201810146C from the US Food and Drug Administration to the Brigham and Women's Hospital (Drs Schneeweiss and Wang). Drs Wang and Schneeweiss were further supported by grants RO1HL141505, RO1AG053302, and RO1AR080194 from the National Institutes of Health.

GUARD CONSORTIUM

Ejemplos de RWE: Trial emulation

		Effect estimates (95% CI)		
			Database study[a]	
Study No.	Trial name	RCT	Adjusted[b]	Crude[b]
1	LEADER	0.87 (0.78 to 0.97)	0.82 (0.76 to 0.87)	0.57 (0.54 to 0.61)
2	DECLARE-TIMI58	0.83 (0.73 to 0.95)	0.69 (0.59 to 0.81)	0.47 (0.41 to 0.53)
3	EMPA-REG	0.86 (0.74 to 0.99)	0.83 (0.73 to 0.95)	0.63 (0.57 to 0.70)
4	CANVAS	0.86 (0.75 to 0.97)	0.77 (0.70 to 0.85)	0.58 (0.54 to 0.62)
5	CARMELINA	1.02 (0.89 to 1.17)	0.90 (0.84 to 0.96)	0.90 (0.86 to 0.95)
6	TECOS	0.98 (0.88 to 1.09)	0.89 (0.86 to 0.91)	0.81 (0.79 to 0.84)
7	SAVOR-TIMI	1.00 (0.89 to 1.12)	0.81 (0.76 to 0.86)	0.65 (0.62 to 0.69)
8	LEAD-2	0 (−0.20 to 0.20)	0.05 (−0.11 to 0.22)	0.01 (−0.11 to 0.13)
9	TRITON-TIMI	0.81 (0.73 to 0.90)	0.88 (0.79 to 0.97)	0.70 (0.65 to 0.76)
10	PLATO	0.84 (0.77 to 0.92)	0.92 (0.83 to 1.02)	0.84 (0.78 to 0.91)

GUARD CONSORTIUM

¡Gracias!