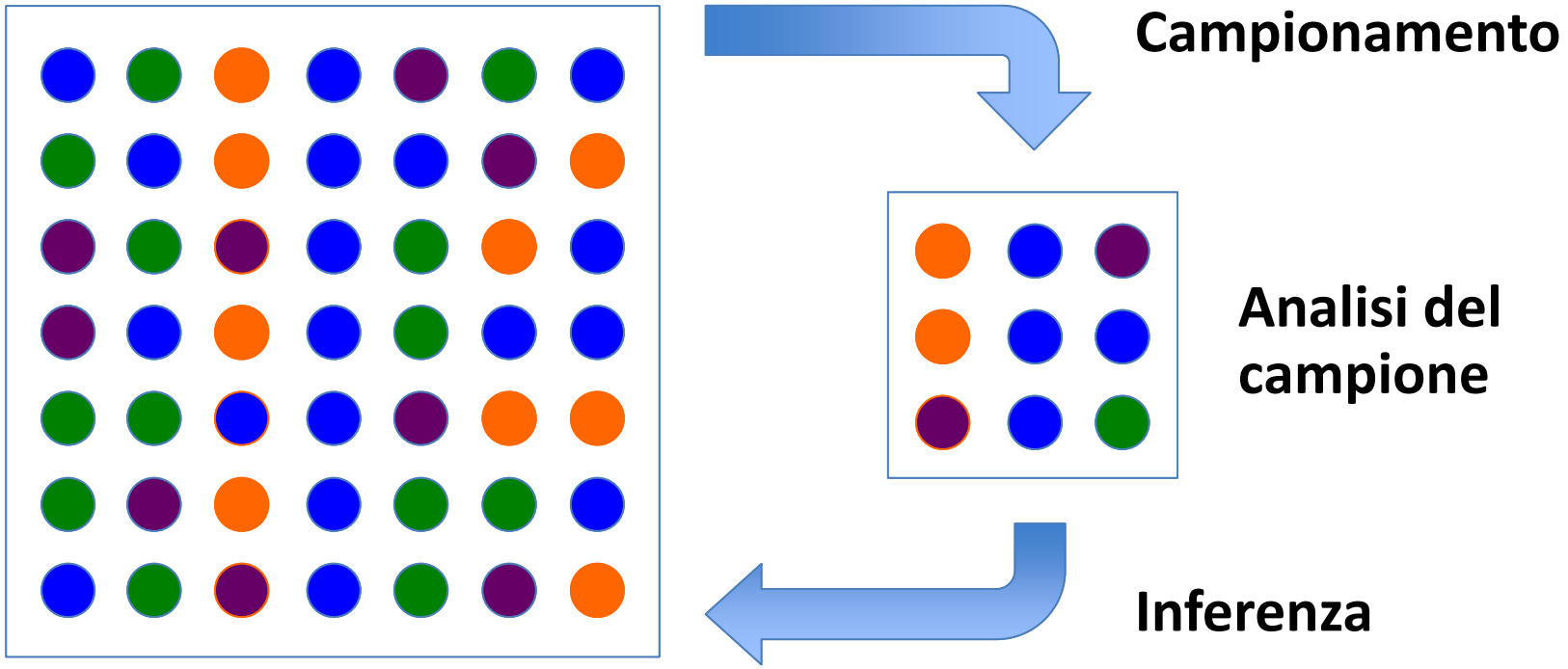
Campione ≠ Popolazione



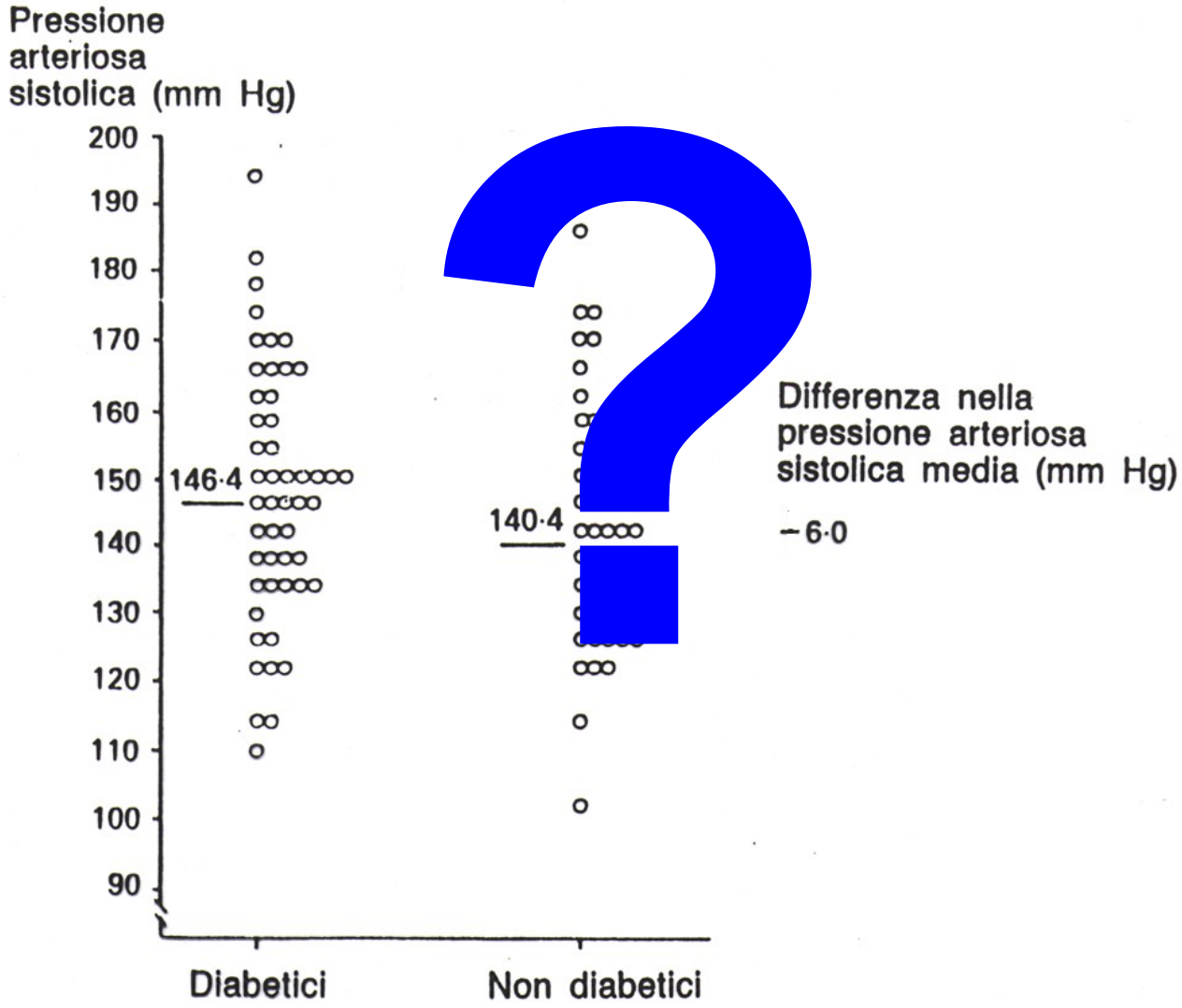
L'INFERENZA STATISTICA SI OCCUPA DI ESTENDERE AD UNA INTERA POPOLAZIONE LE INFORMAZIONI CHE SI RICAVANO DALLE STATISTICHE CAMPIONARIE

VARIABILE DI TIPO QUALITATIVO…

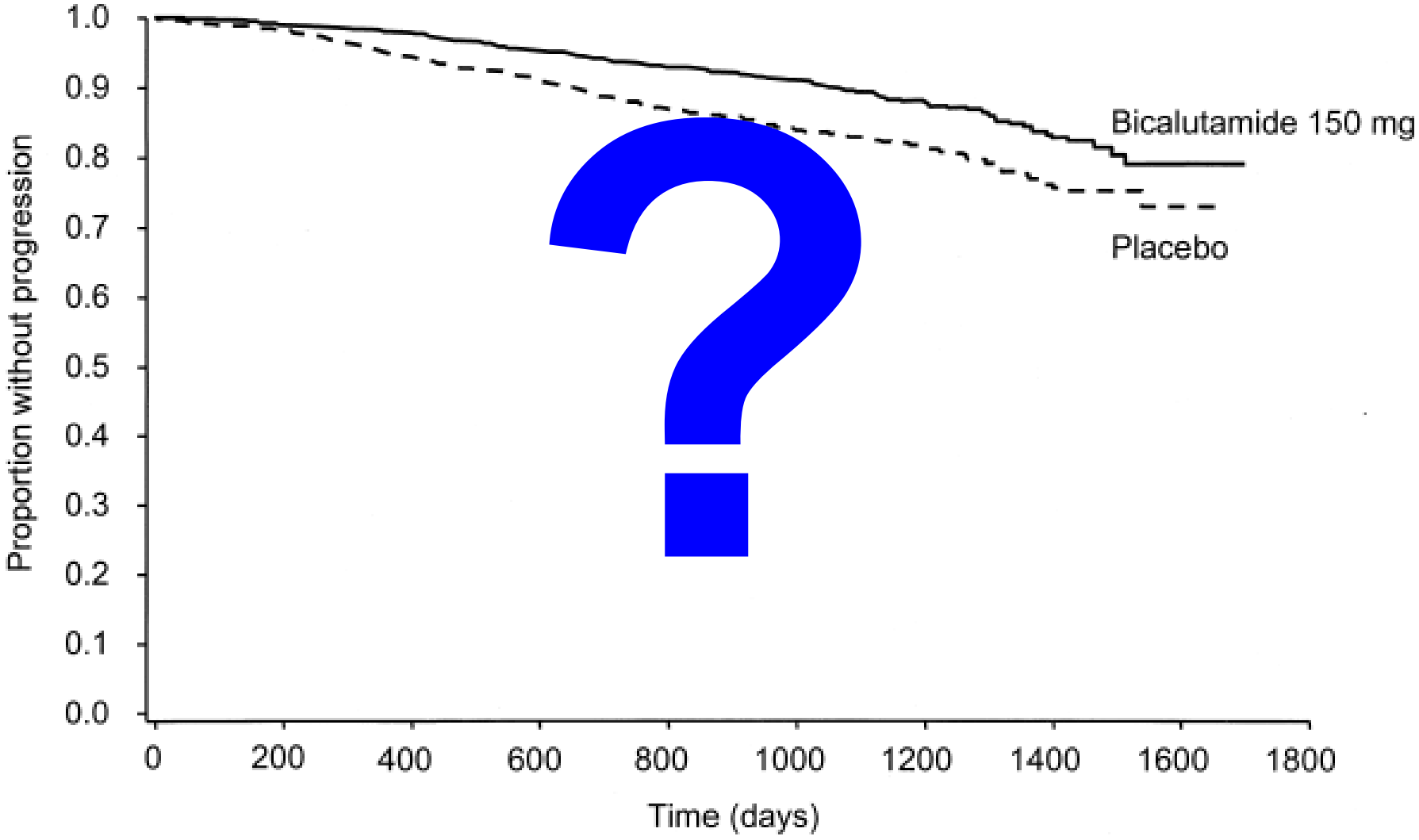
Table 2. Rates of Radiologic Response.*

Subgroup and Variable	Cetuximab plus Irinotecan	Cetuximab
No. of patients	218	111
Response — no. (%)		
Complete response	0	0
Partial response	50 (22.9)	12 (10.8)
Stable disease	71 (32.6)	24 (21.6)
Progressive disease	68 (31.2)	59 (53.2)
Could not be evaluated	29 (13.3)	16 (14.4)
Overall response	50 (22.9	12 (10.8

VARIABILE DI TIPO QUANTITATIVO…



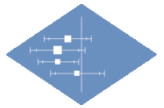
VARIABILE DI TIPO TEMPO A EVENTO…



- **L'effetto osservato è solo uno dei possibili risultati dello studio legati all'assegnazione casuale dei soggetti ai trattamenti.**
- **Se non c'è un reale effetto del trattamento, è comunque possibile che per caso si osservino differenze tra i gruppi.**
- **Se c'è un reale effetto del trattamento, è comunque possibile che per caso non si osservi quella differenza fra i trattamenti.**

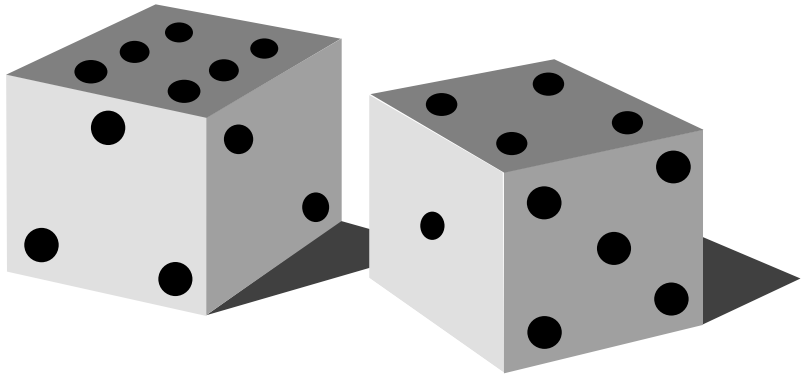
effetto osservato = effetto del trattamento + effetto del caso

(in assenza di errore sistematico!)



IL METODO STATISTICO

- **Il metodo statistico ci permette di valutare quanto può essere affidabile il risultato osservato.**
- **Se i soggetti sono stati assegnati in modo casuale ai trattamenti, si è in grado di calcolare esattamente il ruolo probabile del caso.**



Hypothesis Testing and Jury Trials

In tribunale	Test statistico
Presunzione di innocenza	Ipotesi di non differenza tra i trattamenti (H_0)
Evidenza di colpevolezza	Ipotesi di efficacia del trattamento in esame (H_1)

Susan Hilsenbeck, SABCS 2007 (modif. GLP)

# of Hypothesis

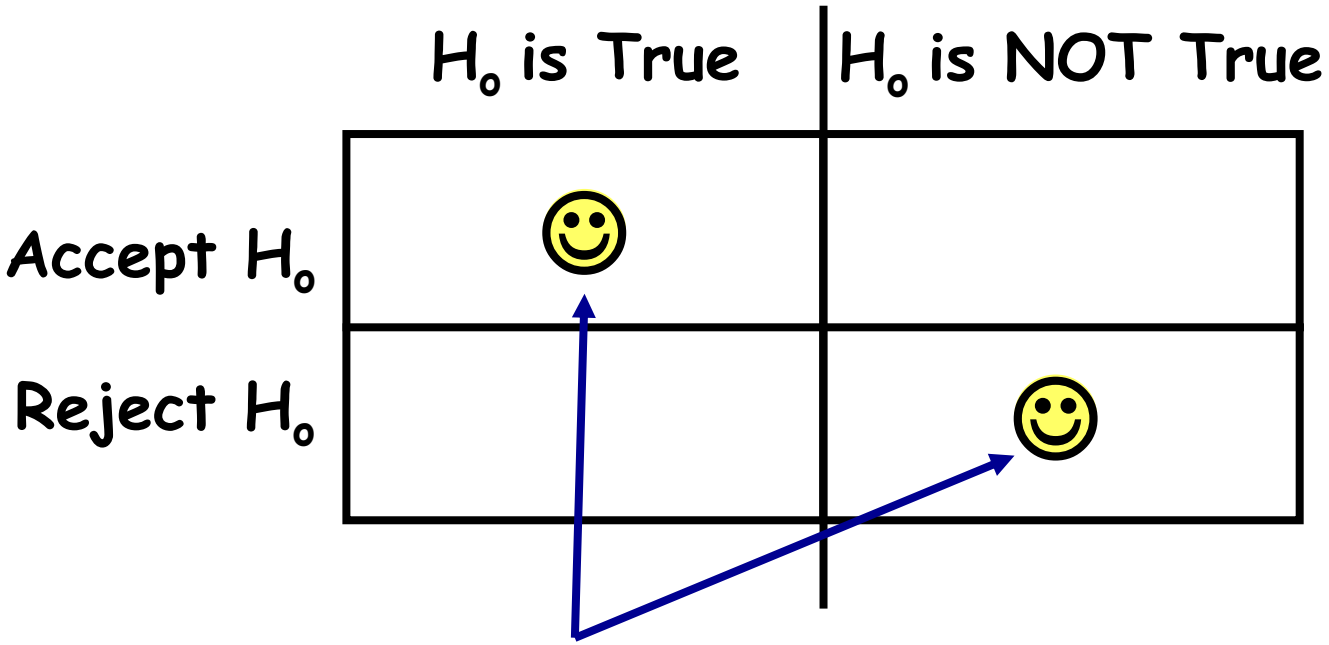
Ma noi non sappiamo quale delle due ipotesi sia quella vera nella realtà (cioè nella popolazione) …

	H_o is True	H_o is NOT True
Accept H_o		
Reject H_o		

Il test statistico può risultare positivo o negativo

Susan Hilsenbeck, SABCS 2007 (modif. GLP)

Making Decisions: Test of Hypothesis



Il test potrebbe quindi dare
la risposta corretta…

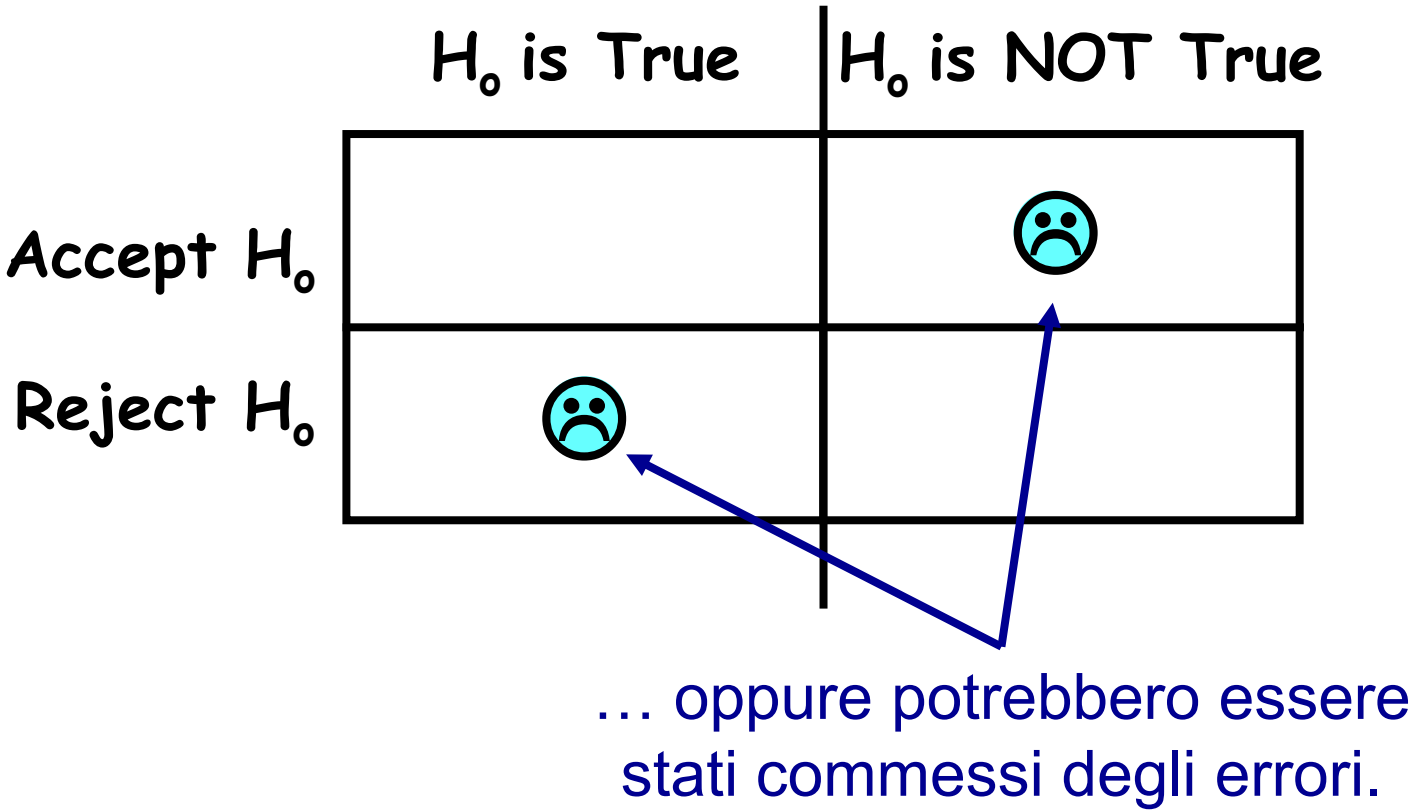
Susan Hilsenbeck, SABCS 2007 (modif. GLP)

Hypothesis Testing and Jury Trials

In tribunale	Test statistico
Presunzione di innocenza	Ipotesi di non differenza tra i trattamenti (H_0)
Evidenza di colpevolezza	Ipotesi di efficacia del trattamento in esame (H_1)
Assolvere l'innocente	Accettazione corretta di H_0
Condannare il colpevole	Corretto rifiuto di H_0
Condannare l'innocente	
Assolvere il colpevole	

Susan Hilsenbeck, SABCS 2007 (modif. GLP)

Making Decisions: Test of Hypothesis

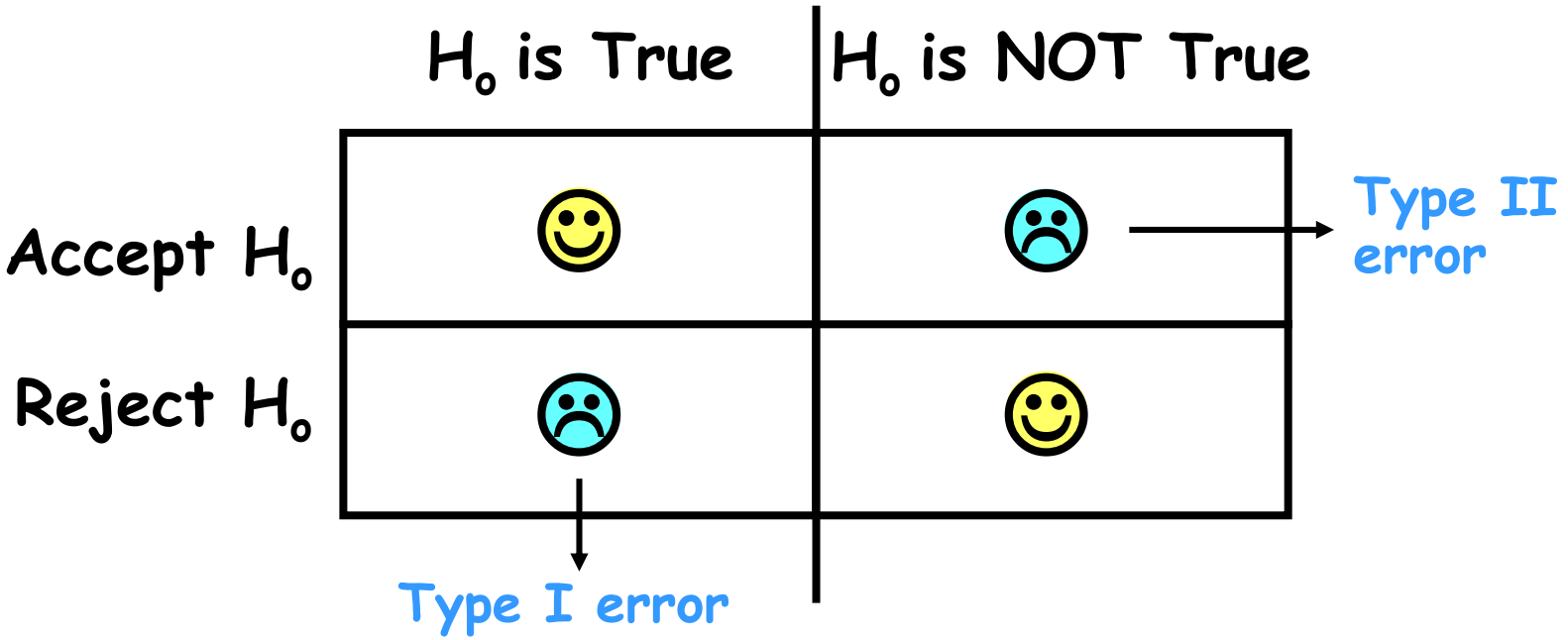


Hypothesis Testing and Jury Trials

In tribunale	Test statistico
Presunzione di innocenza	Ipotesi di non differenza tra i trattamenti (H_0)
Evidenza di colpevolezza	Ipotesi di efficacia del trattamento in esame (H_1)
Assolvere l'innocente	Accettazione corretta di H_0
Condannare il colpevole	Corretto rifiuto di H_0
Condannare l'innocente	Errore di 1° tipo
Assolvere il colpevole	Errore di 2° tipo

Susan Hilsenbeck, SABCS 2007 (modif. GLP)

Making Decisions: Test of Hypothesis



La probabilità di commettere un errore statistico è strettamente dipendente dalle dimensioni del campione

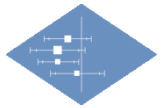
Susan Hilsenbeck, SABCS 2007 (modif. GLP)

ERRORI STATISTICI

- ✓ **Errore di 1° tipo** (errore *alfa*)
  - quando si conclude per un'efficacia del trattamento sperimentale, quando non lo è nella realtà; *lo studio è falsamente positivo*.
- ✓ **Errore di 2° tipo** (errore *beta*)
  - quando si conclude per una non efficacia del trattamento sperimentale, quando invece lo è nella realtà; *lo studio è falsamente negativo*.
- ✓ Il calcolo delle dimensioni del campione mira a contenere la dimensione degli errori statistici entro valori accettabili (5% per l'errore di 1° tipo e 20% per l'errore di 2° tipo)

LIVELLO DI SIGNIFICATIVITA'

- **Rappresenta una soglia:**
  - ✓ ***quando il valore di p risultante dal test di significatività è più piccolo del valore soglia (usualmente α = 0.05), si respinge l'ipotesi nulla in favore di quella alternativa;***
  - ✓ ***se il valore di p è maggiore di α, si accetta l'ipotesi nulla.***
- **Ha solo un valore indicativo, non assoluto.**
- **Non ha niente a che vedere con l'importanza della decisione da prendere o con i costi e le perdite associati agli esiti.**



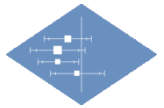
TEST A UNA CODA vs TEST A DUE CODE

- **Il valore di *p a una coda* rappresenta la probabilità che si ha di osservare una differenza uguale o più estrema di quella osservata, *in una sola direzione*, se l'ipotesi nulla è vera:**

ipotesi alternativa: braccio A* > *braccio B

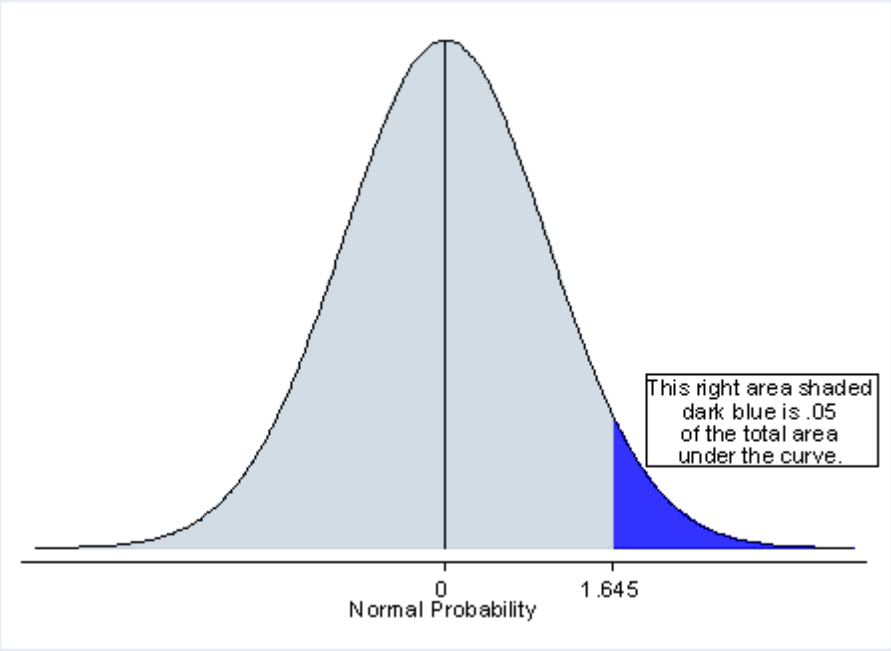
- **il valore di *p a due code* rappresenta la probabilità di rilevare una differenza della stessa entità o più estrema di quella osservata, *in entrambe le direzioni*:**

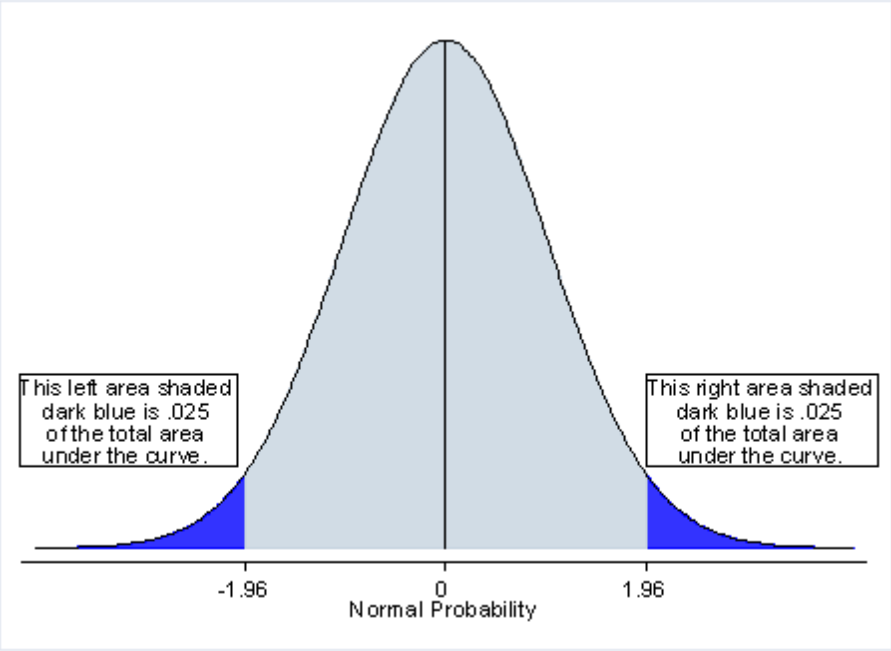
ipotesi alternativa: braccio A* <> *braccio B



When you conduct a test of statistical significance, you are given a p-value somewhere in the output.

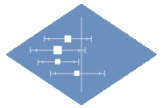
- A one-tailed test allots all of your alpha to testing the statistical significance in the one direction of interest.
- A two-tailed test allots half of your alpha to testing the statistical significance in one direction and half of your alpha to testing statistical significance in the other direction.





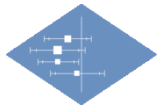
MA *p=0.04* E' COSI' DIVERSO DA *p=0.06*?

- **Il livello di significatività convenzionale ha solo valore indicativo, non assoluto.**
- **Se p=0.05 è la linea di demarcazione tra un risultato significativo e uno non significativo, è però raccomandabile che valori di 0.04 e 0.06 conducano a conclusioni analoghe, non certamente opposte.**



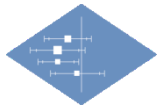
SIGNIFICATIVITA' STATISTICA E RILEVANZA CLINICA

- **Il più delle volte con le parole "*miglioramento significativo*" ci si riferisce al risultato di un test di verifica di ipotesi statistica piuttosto che alla rilevanza clinica del fenomeno osservato.**
- **Il giudizio sulla rilevanza clinica di un fenomeno osservato non può prescindere dalla differenza attesa (auspicata) dallo Sperimentatore in fase di programmazione dell'esperimento.**



SIGNIFICATIVITA' STATISTICA E RILEVANZA CLINICA

- **In presenza di campioni molto ampi, piccole differenze possono risultare statisticamente significative anche se scarsamente rilevanti nella pratica clinica.**
- **Piccoli campioni possono generare grandi differenze, ma determinate in modo così impreciso da non risultare statisticamente significative.**



SINOSSI DEI TEST STATISTICI (1)

	variabili quantitative	
	test parametrici	test non parametrici
due gruppi di trattamento composti da differenti individui	*t*-test per campioni indipendenti	test di *Mann-Whitney*
tre o più gruppi di trattamento composti da differenti individui	analisi della varianza (covarianza)	test di *Kruskall-Wallis*
prima e dopo un singolo trattamento negli stessi individui	*t*-test per dati appaiati	test di *Wilcoxon* per dati appaiati
trattamenti multipli negli stessi individui	analisi della varianza per misure ripetute	test di *Friedman*

DISTRIBUZIONE SIMMETRICA *Vs* NON SIMMETRICA

Quali test statistici?

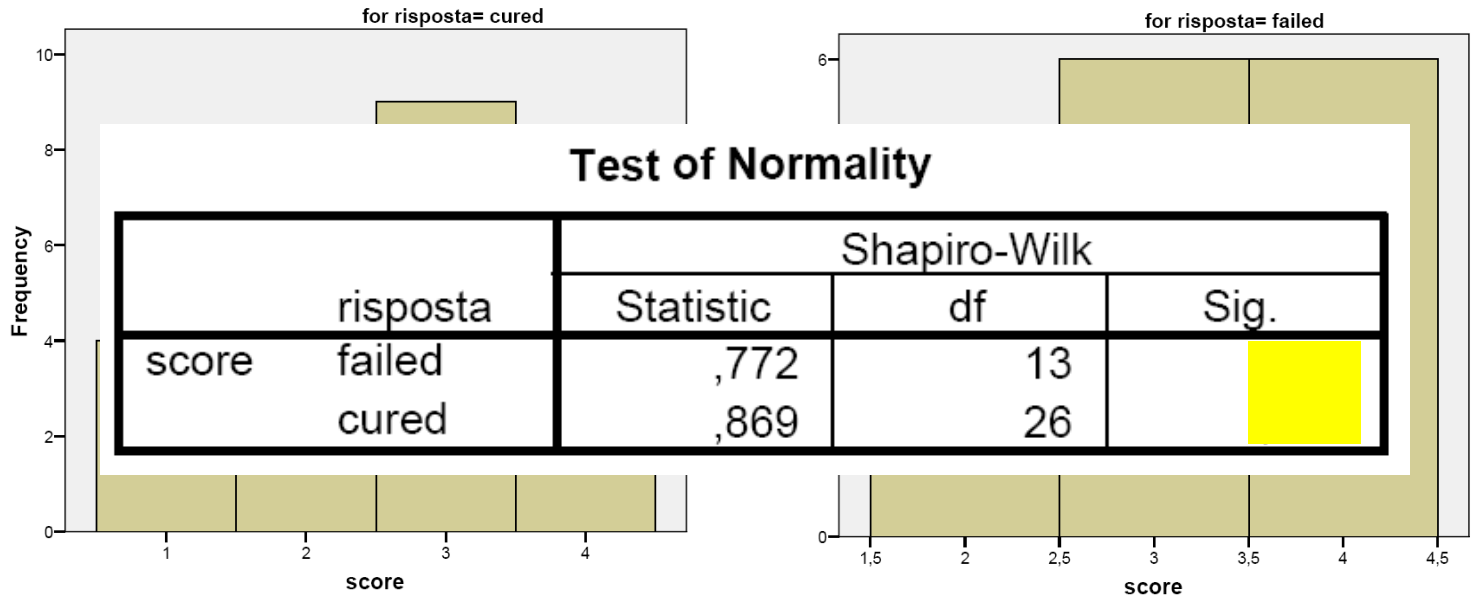
	risposta	N	Mean	Std. Deviation	Std. Error Mean
score	cured	26	2,73	1,041	,204
	failed	13	3,38	,650	,180

	t	df	Sig. (2-tailed)	Mean Difference
score	-2,064	37		-,654

DISTRIBUZIONE SIMMETRICA *Vs* NON SIMMETRICA

Quali test statistici?

	risposta	N	Mean	Std. Deviation	Std. Error Mean
score	cured	26	2,73	1,041	,204
	failed	13	3,38	,650	,180



Test of Normality

		Shapiro-Wilk		
	risposta	Statistic	df	Sig.
score	failed	,772	13	
	cured	,869	26	

DISTRIBUZIONE SIMMETRICA *Vs* NON SIMMETRICA

Quali test statistici?

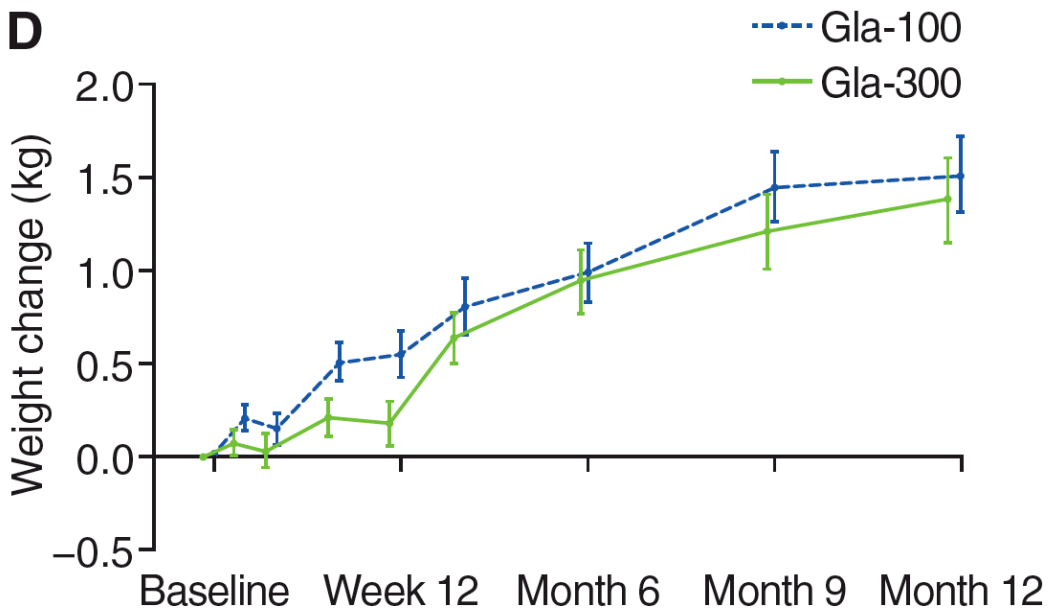
	risposta	N	Mean	Std. Deviation	Std. Error Mean
score	cured	26	2,73	1,041	,204
	failed	13	3,38	,650	,180

	score
Mann-Whitney U	109,000
Asymp. Sig. (2-tailed)	
Exact Sig. [2*(1-tailed Sig.)]	

One-year sustained glycaemic control and less hypoglycaemia with new insulin glargine 300 U/ml compared with 100 U/ml in people with type 2 diabetes using basal plus meal-time insulin: the EDITION 1 12-month randomized trial, including 6-month extension

M. C. Riddle[1], H. Yki-Järvinen[2], G. B. Bolli[3], M. Ziemen[4], I. Muehlen-Bartmer[4], S. Cissokho[5] & P. D. Home[6]

Diabetes, Obesity and Metabolism 2015.



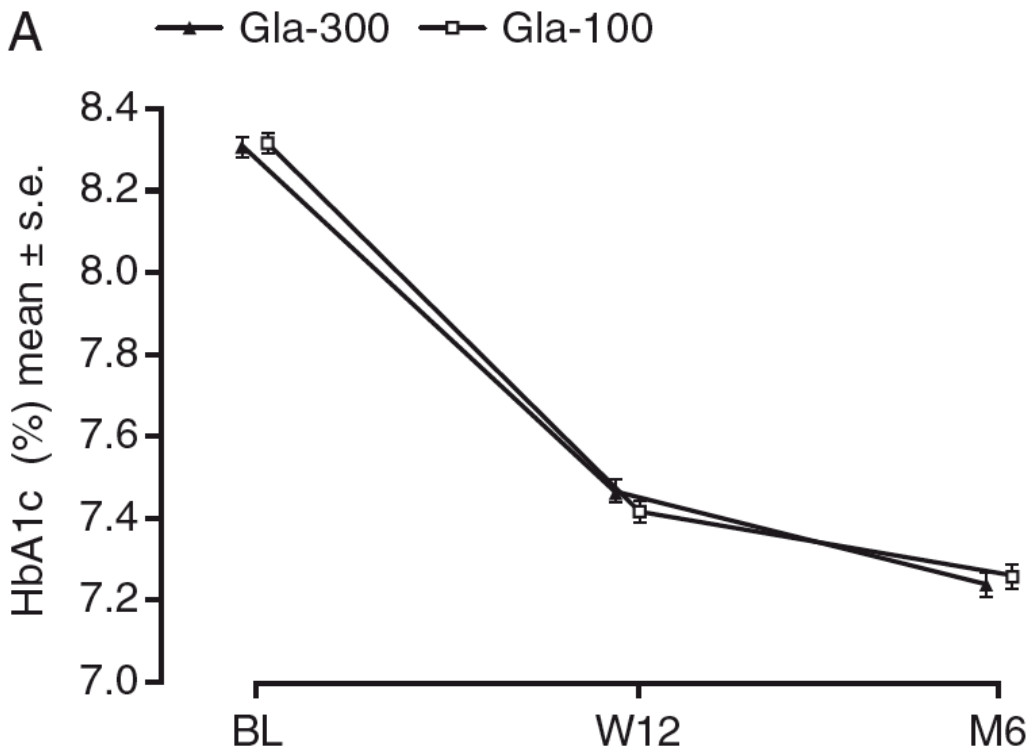
The change in weight frombaseline to month 12 was analysed using an analysis of co-variance model adjust-ed by HbA1c strata and world region.

L'analisi della covarianza è utile quando all'analisi dei dati di una certa variabile, posta sotto controllo in base ad un determinato disegno sperimentale, risulta associata una covariata per cui sia difficile creare dei gruppi omogenei da sottoporre alla sperimentazione

Patient-level meta-analysis of the EDITION 1, 2 and 3 studies: glycaemic control and hypoglycaemia with new insulin glargine 300 U/ml versus glargine 100 U/ml in people with type 2 diabetes

R. Ritzel[1], R. Roussel[2,3,4], G. B. Bolli[5], L. Vinet[6], C. Brulle-Wohlhueter[7], S. Glezer[7] & H. Yki-Järvinen[8]

Diabetes, Obesity and Metabolism 2015.

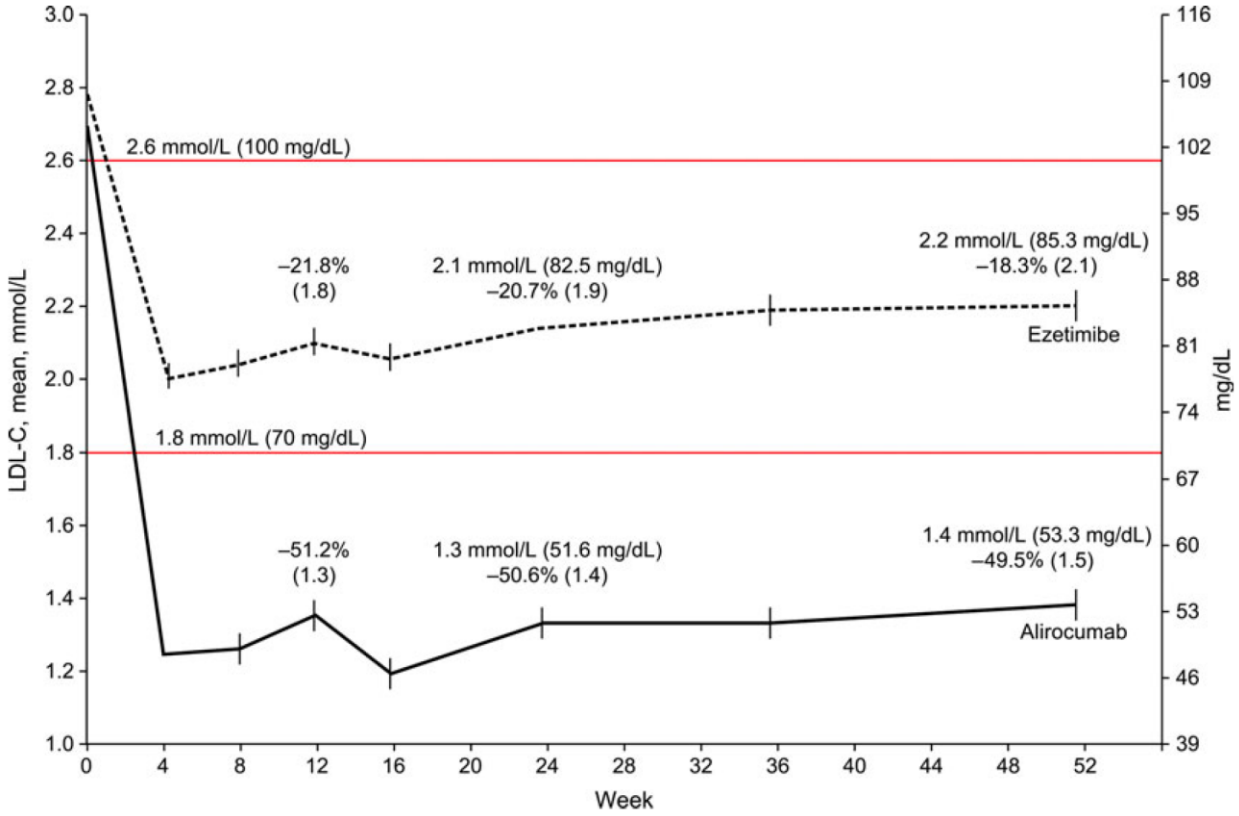


Change in HbA1c, and all other efficacy measures except insulin dose, was analysed using a mixed model for repeated measurements (MMRM).

Efficacy and safety of alirocumab in high cardiovascular risk patients with inadequately controlled hypercholesterolaemia on maximally tolerated doses of statins: the ODYSSEY COMBO II randomized controlled trial

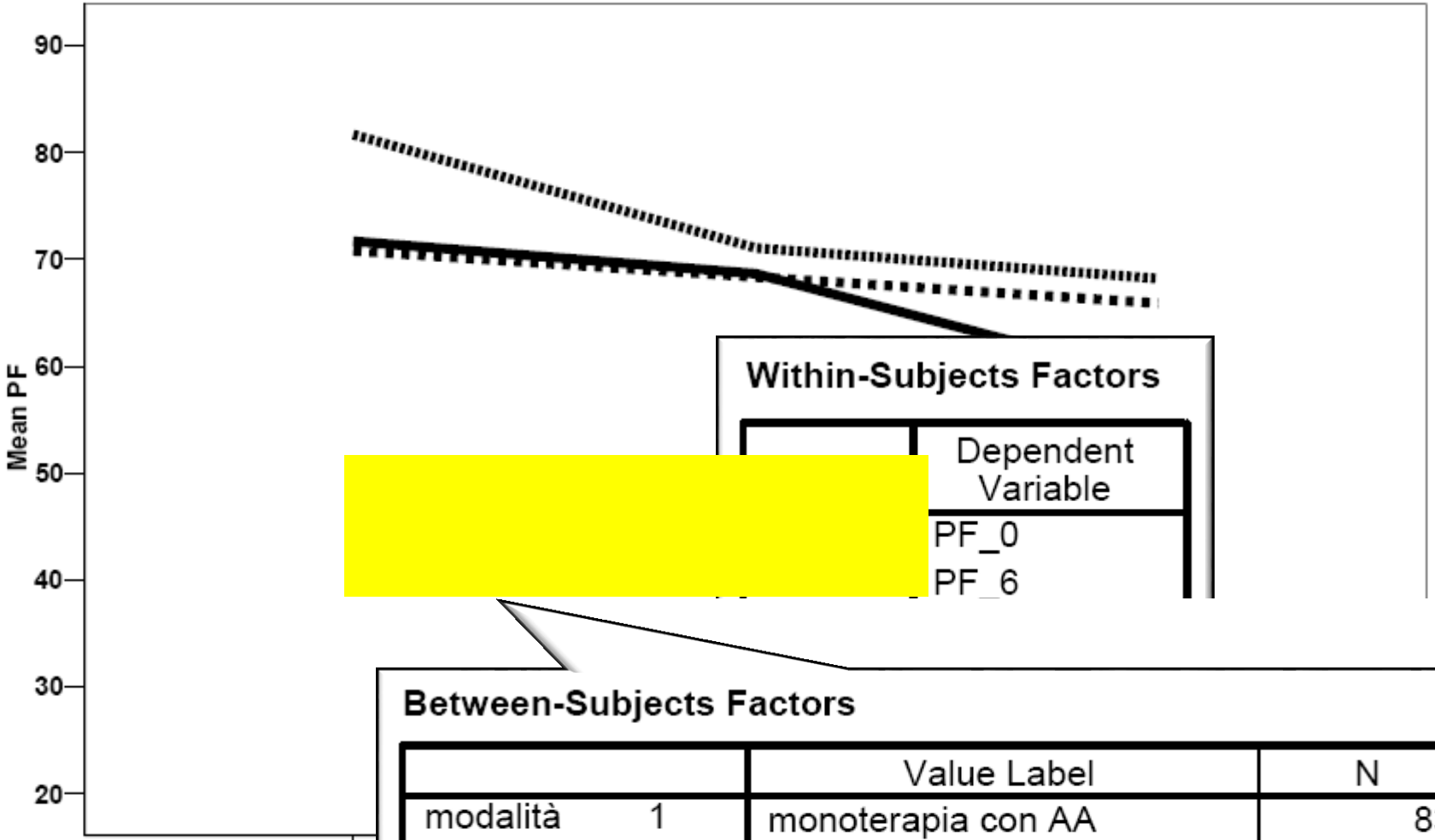
Christopher P. Cannon[1]*, Bertrand Cariou[2], Dirk Blom[3], James M. McKenney[4], Christelle Lorenzato[5], Robert Pordy[6], Umesh Chaudhari[7], and Helen M. Colhoun[8], for the ODYSSEY COMBO II Investigators[†]

European Heart Journal (2015) **36**, 1186–1194



The primary endpoint was analysed using a mixed effect model with a repeated measures (MMRM) approach.

ANOVA PER MISURE RIPETUTE



Within-Subjects Factors

	Dependent Variable
	PF_0
	PF_6

Between-Subjects Factors

		Value Label	N
modalità HT iniziale	1	monoterapia con AA	83
			99
			105

ANOVA PER MISURE RIPETUTE

Tests of Within-Subjects Effects

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
	Sphericity Assumed	14931.830	2	7465.915	37.822	.000
	Huynh-Feldt	14931.830	1.932	7728.519	37.822	.000
	Lower-bound	14931.830	1.000	14931.830	37.822	.000
	Sphericity Assumed	4408.191	4	1102.048	5.583	.000
	Huynh-Feldt	4408.191	3.864	1140.811	5.583	.000
	Lower-bound	4408.191	2.000	2204.095	5.583	.004
Error(tempo)	Sphericity Assumed	112119.798	568	197.394		
	Greenhouse-Geisser	112119.798	541.300	207.131		
	Huynh-Feldt	112119.798	548.700	204.337		
	Lower-bound	112119.798	284.000	394.788		

Tests of Between-Subjects Effects

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	4094216.506	1	4094216.506	4595.793	.000
Error	253004.740	284	890.862		

SINOSSI DEI TEST STATISTICI (2)

	variabili qualitative	variabili del tipo "tempo a evento"
due gruppi di trattamento composti da differenti individui	test chi^2 per tabella di contingenza	*log-rank* test
tre o più gruppi di trattamento composti da differenti individui	test chi^2 per tabella di contingenza	*log-rank* test
prima e dopo un singolo trattamento negli stessi individui	test di *McNemar*	-
trattamenti multipli negli stessi individui	-	-

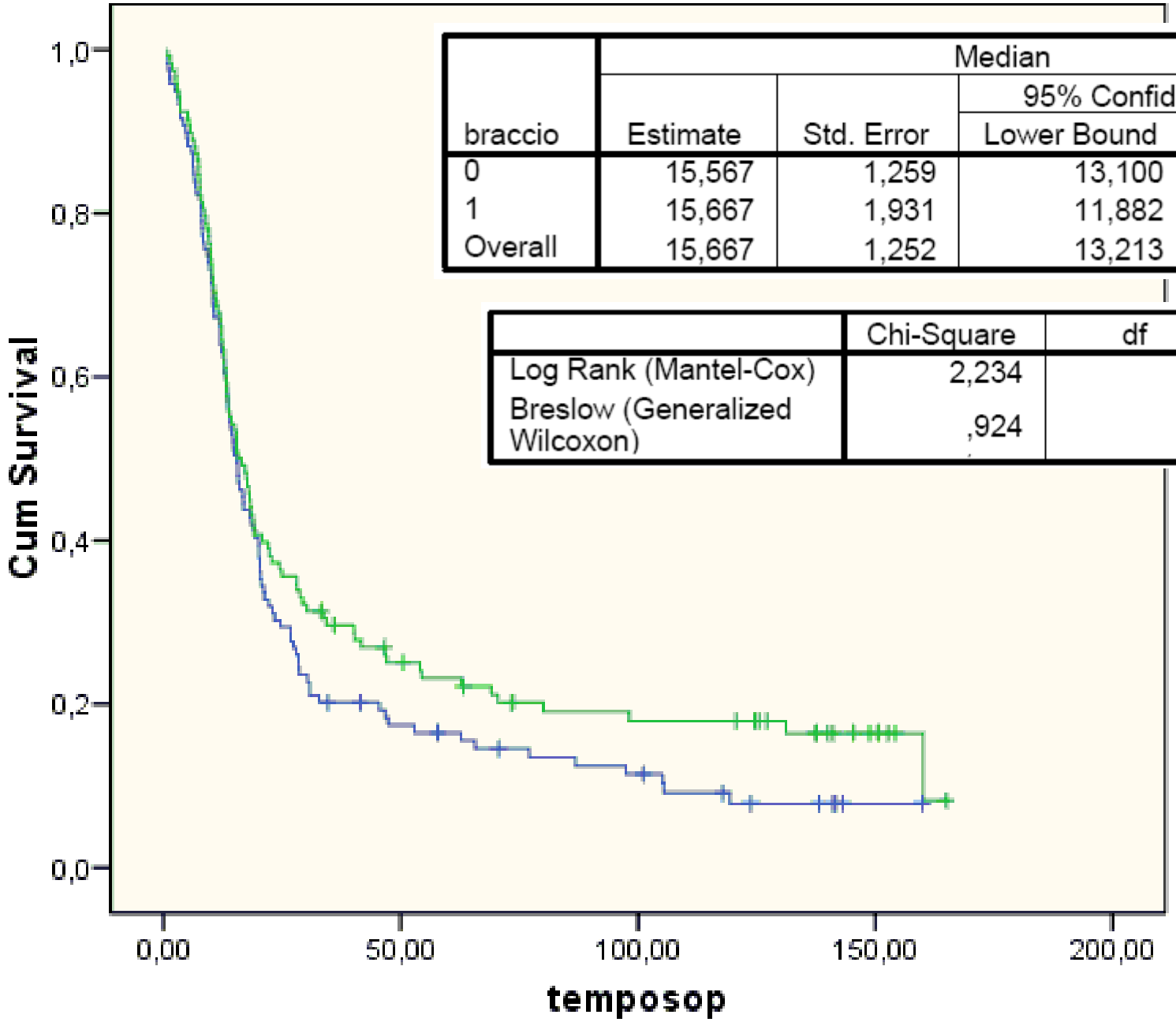
TABELLE DI CONTINGENZA

Quali test statistici?

			risposta		Total
			non risposta	risposta	
trattamento	test	Count	20	8	28
		Expected Count	23,1		
	control	Count	32	3	35
		Expected Count	28,9	6,1	
Total		Count	52	11	63

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
	[b]	1			,040
		1			
					,040
N of Valid Cases	63				

a. Computed only for a 2x2 table



braccio	Median			
			95% Confidence Interval	
	Estimate	Std. Error	Lower Bound	Upper Bound
0	15,567	1,259	13,100	18,034
1	15,667	1,931	11,882	19,452
Overall	15,667	1,252	13,213	18,121

	Chi-Square	df	Sig.
Log Rank (Mantel-Cox)	2,234	1	,135
Breslow (Generalized Wilcoxon)	,924	1	,336

L'ANALISI MULTIVARIATA

Descrive la relazione tra una variabile di outcome (v. dipendente) e i suoi diversi determinanti (v. indipendenti) con lo scopo di predire l'outcome in condizioni diverse

Il principale vantaggio consiste nel poter apprezzare – simultaneamente e indipendentemente – gli effetti di tutte le variabili considerate, depurandoli dalle reciproche interferenze

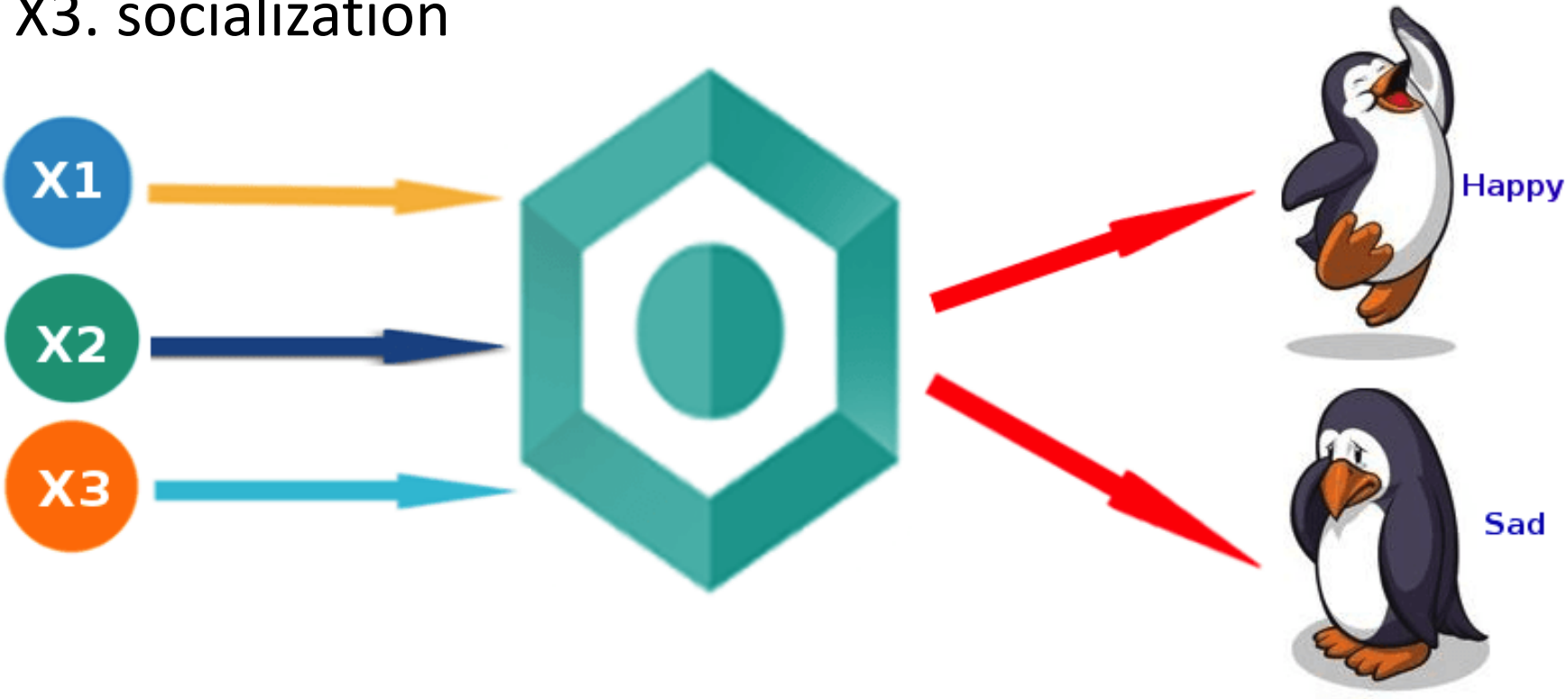
MULTIVARIATE ANALYSIS

Independent variables

X1. food
X2. water temperature
X3. socialization

Dependent variable

Penguin mood



L'ANALISI MULTIVARIATA

Modelli di analisi p

outcome (v. dipend

Poisson regression may also be appropriate for rate data, where the rate is a count of events divided by some measure of that unit's exposure (i.e ARR in MS)

Regressione Lineare

qualitative

Regressione Logistica **quantitative** **qualitativa**

qualitative

L'ANALISI MULTIVARIATA

Modelli di analisi più comuni, per tipo di variabile di outcome (v. dipendente)

	variabili indipendenti	variabile dipendente
Regressione Lineare	quantitative qualitative	quantitativa
Regressione Logistica	quantitative qualitative	qualitativa
Cox's Regression	quantitative qualitative	tempo a evento

Effect of alirocumab, a monoclonal antibody to PCSK9, on long-term cardiovascular outcomes following acute coronary syndromes: Rationale and design of the ODYSSEY Outcomes trial

Gregory G. Schwartz, MD, PhD,[a,q] Laurence Bessac, MD,[b,c] Lisa G. Berdan, PA, MHS,[d] Deepak L. Bhatt, MD, MPH,[e] Vera Bittner, MD,[f] Rafael Diaz, MD,[g] Shaun G. Goodman, MD, MSc,[h] Corinne Hanotin, MD,[b,c] Robert A. Harrington, MD,[i] J. Wouter Jukema, MD, PhD,[j] Kenneth W. Mahaffey, MD,[i] Angèle Moryusef, MD,[b,c] Robert Pordy, MD,[k] Matthew T. Roe, MD, MPH,[d] Tyrus Rorick, RN,[d] William J. Sasiela, PhD,[k] Cheerag Shirodaria, MBBS,[l] Michael Szarek, PhD,[m] Jean-François Tamby, MD,[b,c] Pierluigi Tricoci, MD,[d] Harvey White, MBBS, DSc,[n] Andreas Zeiher, MD,[o] and Philippe Gabriel Steg, MD[p,q]

(Am Heart J 2014;168:682-689.e1.)

Statistical analysis

The primary outcome will be analyzed with the log-rank test procedure stratified by geographical region. The effect of the time from ACS event to randomization will be assessed using a Cox proportional hazards model including the time from ACS event as a covariate, the treatment group, and the interaction.

ERRORI STATISTICI

- ✓ **Errore di 1° tipo** (errore *alfa*)
  - quando si conclude per un'efficacia del trattamento sperimentale, quando non lo è nella realtà; *lo studio è falsamente positivo*.
- ✓ **Errore di 2° tipo** (errore *beta*)
  - quando si conclude per una non efficacia del trattamento sperimentale, quando invece lo è nella realtà; *lo studio è falsamente negativo*.
- ✓ Il calcolo delle dimensioni del campione mira a contenere la dimensione degli errori statistici entro valori accettabili (5% per l'errore di 1° tipo e 20% per l'errore di 2° tipo)

DIMENSIONAMENTO CAMPIONARIO

Punti caratterizzanti:

- ✓ definizione del **livello basale di efficacia** (stima dell'effetto nel gruppo di controllo)
- ✓ **beneficio minimo** di interesse clinico (a favore del trattamento sperimentale) che si intende dimostrare

DIMENSIONAMENTO CAMPIONARIO

Punti caratterizzanti:

- ✓ definizione del **livello basale di efficacia** (stima dell'effetto nel gruppo di controllo)
- ✓ **beneficio minimo** di interesse clinico (a favore del trattamento sperimentale) che si intende dimostrare
- ✓ dimensione (accettabile) dell'errore di 1° tipo = **livello di significatività statistica**
  - quando il valore di p risultante dal test di significatività è più piccolo del valore soglia (usualmente 5%), si considera lo studio positivo;
  - se il valore di p è maggiore del 5%, si considera lo studio negativo
  - non ha niente a che vedere con l'importanza clinica dell'effetto osservato
- ✓ **potenza statistica** (1 *meno* errore di 2° tipo) che si intende perseguire
  - usualmente fissata all'80%

Superiorità Vs Non-inferiorità

Si ritiene che il trattamento in esame "A" abbia le potenzialità per migliorare il trattamento standard "B" almeno di una quantità Δ

studio di superiorità

A > B di una quantità Δ di interesse clinico

studio di non inferiorità

A < B non oltre una quantità M di rilevanza clinica

Superiorità Vs Non-inferiorità

La dimensione di Δ dipende da:

- ✓ **tipo e fase della malattia in atto**
- ✓ **efficacia dei trattamenti disponibili**
- ✓ **caratteristiche del trattamento sperimentale**

studio di superiorità

A > B di una quantità Δ di interesse clinico

studio di non inferiorità

A < B non oltre una quantità M di rilevanza clinica

When Are "Positive" Clinical Trials in Oncology Truly Positive?

Alberto Ocana, Ian F. Tannock

J Natl Cancer Inst 2011;103:16–20

We would define a positive trial as one in which the predefined value of *delta* represents a clinically important difference in an endpoint that directly reflects benefit (mainly OS or quality of life) to patients and for which the results provide a best estimate of the

Consistent with a recent commentary suggesting the need to increase the value of *delta* in future clinical trials, we provide an estimate of *delta* that would be generally accepted as representing a minimum clinically important difference in the primary endpoint: approximately 3 months increase in median OS for patients with advanced metastatic solid tumors (usually corresponding to an

DIMENSIONAMENTO CAMPIONARIO PER VARIABILI DI TIPO QUANTITATIVO:

- **Per una variabile di risposta di tipo quantitativo, occorre precisare:**
  - **media e deviazione standard del gruppo di controllo;**
  - **differenza di interesse clinico;**
  - **l'errore di 1° tipo;**
  - **la potenza del test statistico.**

Effect of a monoclonal antibody to PCSK9, REGN727/ SAR236553, to reduce low-density lipoprotein cholesterol in patients with heterozygous familial hypercholesterolaemia on stable statin dose with or without ezetimibe therapy: a randomised controlled trial

Evan A Stein, Dan Gipe, Jean Bergeron, Daniel Gaudet, Robert Weiss, Robert Dufour, Richard Wu, Robert Pordy

Lancet 2012; 380: 29–36

Sample size and power calculations were based on the ability to detect a treatment difference in LDL-C of from baseline to week 12 with a standard deviation range of 20–30% from the completed phase 1 studies.[14]

Monotherapy with the PCSK9 inhibitor alirocumab versus ezetimibe in patients with hypercholesterolemia: Results of a 24 week, double-blind, randomized trial

Eli M. Roth [a,*], Marja-Riitta Taskinen [b], Henry N. Ginsberg [c], John J.P. Kastelein [d], Helen M. Colhoun [e], Jennifer G. Robinson [f], Laurence Merlet [g], Robert Pordy [h], Marie T. Baccara-Dinet [i]

International Journal of Cardiology 176 (2014) 55–61

A sample size of 45 patients per treatment arm was calculated to have 95% power to detect a mean difference between alirocumab and ezetimibe of in LDL-C percent change from baseline to week 24 using a 2-sided *t*-test with 5% significance, assuming a common standard deviation (SD) of 25% based on a previous alirocumab trial [1] and with an expected rate of exclusion of 5%.

The Minimal Clinical Interesting Difference (M.C.I.D.)

- it's easily understood by clinicians as a key concept in the interpretability of PRO scores;
- will inform judgments about the success-fulness of an intervention;
- an individual patient achieving the score equal or greater than the MCID might be considered a beneficiary of the intervention, what would lead to the definition of a **responder**.

Validation of the Multiple Sclerosis International Quality of Life questionnaire

MC Simeoni[1], P Auquier[1], O Fernandez[2], P Flachenecker[3,4], S Stecchi[5], CS Constantinescu[6], E Idiman[7], A Boyko[8], AG Beiske[9], T Vollmer[10], N Triantafyllou[11], P O'Connor[12], Y Barak[13], L Biermann[14†], E Cristiano[15], S Atweh[16], DL Patrick[17], S Robitail[1], N Ammoury[18], A Beresniak[19] and J Pelletier[20] on behalf of the MusiQoL study group

Multiple Sclerosis 2008; **14**: 219–230

	Improved patients self-report (1)		Improved patients physician report (2)	
MusiQol dimensions	Delta	ES	Delta	ES
Activity of daily living	6.33***	0.27	7.63***	0.30
PWB	7.36***	0.36	8.8***	0.37
Relationships with friends	−0.08	0.00	−1.6	−0.07
Symptoms	3.89**	0.18	3.11	0.17
Relationships with family	0.32	0.02	−2.63	−0.11
Relationships with health care system	1.09	0.06	2.01	0.08
Sentimental and sexual life	0.76	0.02	4.61	0.16
Coping	4.63***	0.17	4.59	0.14
Rejection	2.61*	0.12	7.25**	0.33
Index	2.90***	0.22	5.23*	0.41

*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

Delta = HRQL score at D21 – HRQL score at inclusion. Effect size = (HRQL score at D21 – HRQL score at inclusion)/standard deviation of HRQL score at inclusion.

(1) Improvement of health status self-reported by the patients.

(2) Improvement of patients' health status rated by the physician on an item assessing clinical global impression.

An Effect Size (ES) of at least 0.2 is recommended as the standard for supporting sensitivity to change

Estimating a minimal clinically important difference for the EuroQol 5-dimension health status index in persons with multiple sclerosis

Christine G Kohn[1,2], Matthew F Sidovar[3], Kirandeep Kaur[1], Yungfen Zhu[1] and Craig I Coleman[1,2*]

Health and Quality of Life Outcomes 2014, **12**:66

In conclusion, the MCID estimate calculated in this study can aid researchers and clinicians when discriminating between patient groups for EQ-5D index scores of PwMS. Our MCID range of 0.050-0.084 for EQ-5D was within the range of MCID estimates of other disease states. In general, patients who have severe disability had higher MCIDs than patients who had mild-moderate disability.

VARIABILE DI RISPOSTA

- di tipo quantitativo
  - assume uno spettro continuo di valori e viene misurata in riferimento a una scala a intervalli costanti.
- di tipo **qualitativo**
  - esprime categorie di risposta del tipo successo / insuccesso (di un trattamento somministrato).
- del tipo “tempo a evento”
  - rappresenta il tempo trascorso fino al verificarsi (o meno) di un evento.

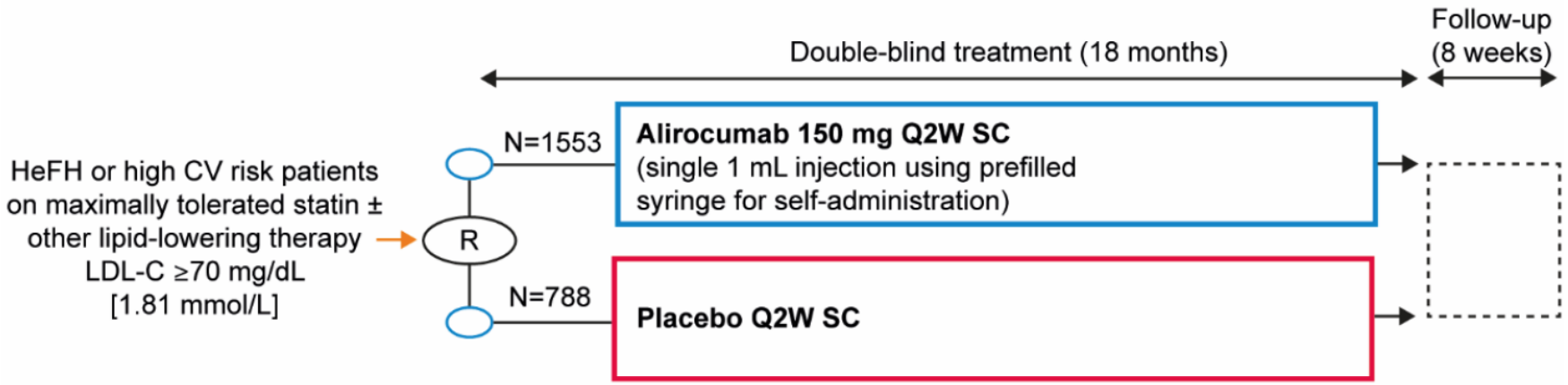
DIMENSIONAMENTO CAMPIONARIO PER VARIABILI DI TIPO QUALITATIVO:

- **Per una variabile di risposta di tipo qualitativo, occorre precisare:**
  - **frequenza di "successi" nel gruppo di controllo;**
  - **differenza di interesse clinico;**
  - **l'errore di 1° tipo;**
  - **la potenza del test statistico.**

Efficacy and Safety of Alirocumab in Reducing Lipids and Cardiovascular Events

Jennifer G. Robinson, M.D., M.P.H., Michel Farnier, M.D., Ph.D., Michel Krempf, M.D., Jean Bergeron, M.D., Gérald Luc, M.D., Maurizio Averna, M.D., Erik S. Stroes, M.D., Ph.D., Gisle Langslet, M.D., Frederick J. Raal, M.D., Ph.D., Mahfouz El Shahawy, M.D., Michael J. Koren, M.D., Norman E. Lepor, M.D., Christelle Lorenzato, M.Sc., Robert Pordy, M.D., Umesh Chaudhari, M.D., and John J.P. Kastelein, M.D., Ph.D., for the ODYSSEY LONG TERM Investigators*

N Engl J Med 2015;372:1489-99.



13.1 DETERMINATION OF SAMPLE SIZE

a sample size of 2100 patients (randomization ratio 2:1, ie, SAR236553: 1400 and placebo: 700) will allow to have long-term safety data in a broad database (at least 1000 patients exposed to SAR236553 for a minimum of 12 months, of which approximately 900 patients exposed to SAR236553 for 18 months).

Moreover, a sample size of 1400 patients treated with SAR236553 will with 95% confidence in the SAR236553 group.

High-Dose Atorvastatin vs Usual-Dose Simvastatin for Secondary Prevention After Myocardial Infarction

The IDEAL Study: A Randomized Controlled Trial

Terje R. Pedersen, MD, PhD
Ole Faergeman, MD, DMSc
John J. P. Kastelein, MD, PhD
Anders G. Olsson, MD, PhD
Matti J. Tikkanen, MD, PhD
Ingar Holme, PhD
Mogens Lytken Larsen, MD, DMSc
Fredrik S. Bendiksen, MD
Christina Lindahl, MD
Michael Szarek, MS
John Tsai, MD

JAMA. 2005;294:2437-2445

The primary clinical outcome was time to first occurrence of a major coronary event, defined as coronary death, hospitalization for nonfatal acute myocardial infarction, or cardiac arrest with resuscitation.

The trial was designed to have 90% power to detect an anticipated 21% relative risk reduction (from 10% to 7.9%) in the primary outcome variable with atorvastatin over 5 years using a 2-tailed α level of .05.

Randomized Trial of Oral Teriflunomide for Relapsing Multiple Sclerosis

Paul O'Connor, M.D., Jerry S. Wolinsky, M.D., Christian Confavreux, M.D., Giancarlo Comi, M.D., Ludwig Kappos, M.D., Tomas P. Olsson, M.D., Ph.D., Hadj Benzerdjeb, M.D., Philippe Truffinet, M.D., Lin Wang, Ph.D., Aaron Miller, M.D., and Mark S. Freedman, M.D., for the TEMSO Trial Group*

N Engl J Med 2011;365:1293-303

A sample of 360 randomly assigned patients per group was required to provide 95% statistical power to detect relative risk reductions of 25% in the annualized relapse rate after 2 years, assuming an annualized relapse rate of 0.74 for the group receiving placebo and a standard deviation of 0.626.

VARIABILE DI RISPOSTA

- di tipo quantitativo
  - assume uno spettro continuo di valori e viene misurata in riferimento a una scala a intervalli costanti.
- di tipo qualitativo
  - esprime categorie di risposta del tipo successo / insuccesso (di un trattamento somministrato).
- del tipo "**tempo a evento**"
  - rappresenta il tempo trascorso fino al verificarsi (o meno) di un evento.

DIMENSIONAMENTO CAMPIONARIO PER VARIABILI TEMPO-A-EVENTO:

- **Per una variabile di risposta di tipo "tempo a evento", occorre precisare:**
  - **stima di sopravvivenza al tempo *t* nel braccio di controllo;**
  - **differenza di interesse clinico;**
  - **l'errore di 1° tipo;**
  - **la potenza del test statistico;**
  - **durata di arruolamento e follow-up;**
  - **max % di pazienti persi al follow-up.**

High-Dose Atorvastatin vs Usual-Dose Simvastatin for Secondary Prevention After Myocardial Infarction

The IDEAL Study: A Randomized Controlled Trial

Terje R. Pedersen, MD, PhD
Ole Faergeman, MD, DMSc
John J. P. Kastelein, MD, PhD
Anders G. Olsson, MD, PhD
Matti J. Tikkanen, MD, PhD
Ingar Holme, PhD
Mogens Lytken Larsen, MD, DMSc
Fredrik S. Bendiksen, MD
Christina Lindahl, MD
Michael Szarek, MS
John Tsai, MD

JAMA. 2005;294:2437-2445

The primary clinical outcome was time to first occurrence of a major coronary event, defined as coronary death, hospitalization for nonfatal acute myocardial infarction, or cardiac arrest with resuscitation.

The trial was designed to have 90% power to detect an anticipated 21% relative risk reduction (from 10% to 7.9%) in the primary outcome variable with atorvastatin over 5 years using a 2-tailed α level of .05.

eriorità

Vista la migliore tollerabilità del trattamento in esame “A”, si è disposti ad accettarne una eventuale minore efficacia rispetto al trattamento standard “B” purché questa non vada oltre un margine M

studio di superiorità

A > B di una quantità Δ di interesse clinico

studio di non inferiorità

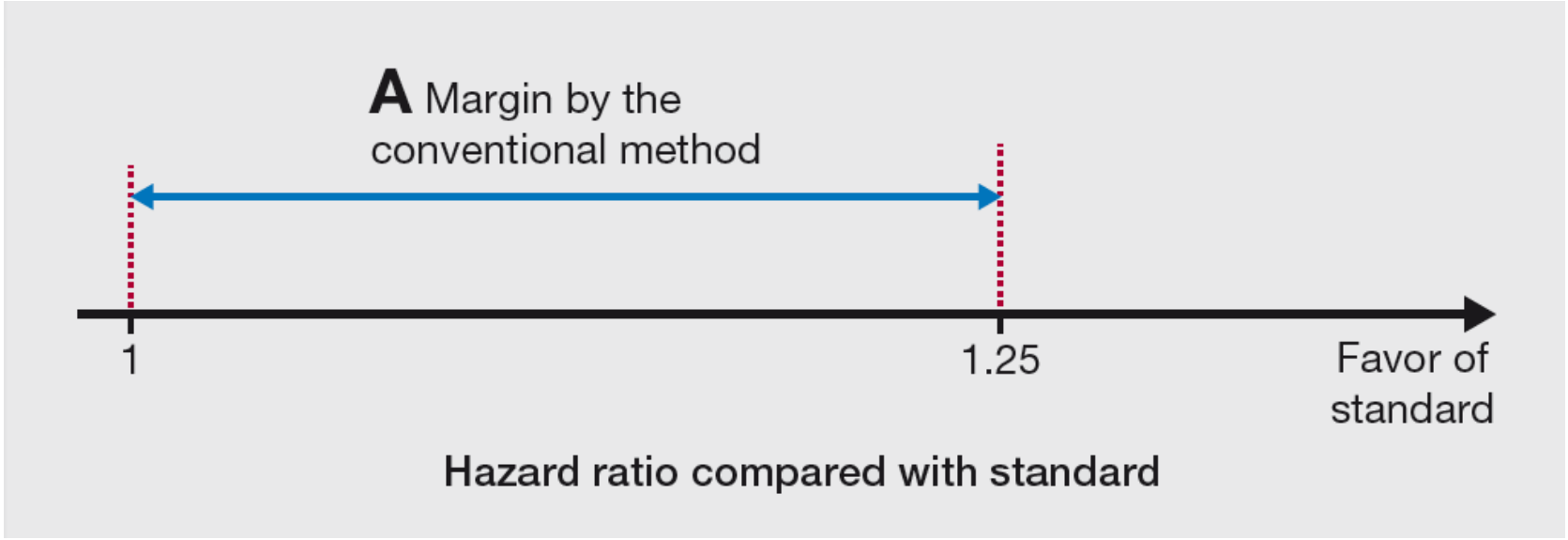
A < B non oltre una quantità M di rilevanza clinica

Statistical Issues and Recommendations for Noninferiority Trials in Oncology: A Systematic Review

Shiro Tanaka[1], Yousuke Kinjo[2], Yoshiki Kataoka[2], Kenichi Yoshimura[1], and Satoshi Teramukai[1]

Clin Cancer Res; 18(7); 1837–47. ©2012 AACR.

Fixed Margin Method



European Medicines Agency
Pre-authorisation Evaluation of Medicines for Human Use

London, 27 July 2005

Doc. Ref. EMEA/CPMP/EWP/2158/99

COMMITTEE FOR MEDICINAL PRODUCTS FOR HUMAN USE (CHMP)

GUIDELINE ON THE CHOICE OF THE NON-INFERIORITY MARGIN

In order to demonstrate non-inferiority, the recommended approach is to pre-specify a margin of non-inferiority in the protocol.

After study completion, a two-sided 95% confidence interval (or one-sided 97.5% interval) for the true difference between the two agents will be constructed. This interval should lie entirely on the positive side of the non-inferiority margin.

It always needs to be tailored specifically to the particular clinical context and

Intensive versus Moderate Lipid Lowering with Statins after Acute Coronary Syndromes

Christopher P. Cannon, M.D., Eugene Braunwald, M.D., Carolyn H. McCabe, B.S., Daniel J. Rader, M.D., Jean L. Rouleau, M.D., Rene Belder, M.D., Steven V. Joyal, M.D., Karen A. Hill, B.A., Marc A. Pfeffer, M.D., Ph.D., and Allan M. Skene, Ph.D., for the Pravastatin or Atorvastatin Evaluation and Infection Therapy–Thrombolysis in Myocardial Infarction 22 Investigators*

N Engl J Med 2004;350:1495-504.

The study was designed to establish the noninferiority of pravastatin as compared with atorvastatin with respect to the time to an end-point event.

For the comparison of pravastatin with atorvastatin, we defined the prespecified boundary for noninferiority as an upper limit of the one-sided 95 percent confidence interval of the relative risk at two years of less than 1.17 (corresponding to a hazard ratio throughout follow-up of 1.198).

Highlights from the Tenth World Conference on Lung Cancer

TRACEY L. EVANS

The University of Pennsylvania Medical Center—Presbyterian, Division of Hematology/Oncology, Philadelphia, Pennsylvania, USA

The Oncologist 2004;9:232-238

	Pemetrexed (n = 283)	Docetaxel (n = 288)
Median survival	8.3 months	7.9 months
Hazard ratio* (95% CI)	0.99 (0.8-1.2)	
1-year survival rate	29.7%	29.7%

*This study was initially designed to be a noninferiority trial using the fixed margin method. Using this statistical test, pemetrexed would be declared no worse than 10% less efficacious than docetaxel if the upper limit of the 95% CI of the overall survival hazard ratio was ≤1.11. The upper limit of the hazard ratio observed in the study was 1.2. By another retrospectively applied test of noninferiority, the percent retention method, the comparator must demonstrate preservation of no less than 50% of the benefit of the standard arm. Pemetrexed in this study had a 102% retention of efficacy (95% CI = 52%-157%, p value = 0.047). Therefore, by this test, pemetrexed would be declared noninferior.

Statistical Review Addendum #1

Medical Division: Oncology Drug Products (HFD-150)
Biometrics Division: Division of Biometrics I (HFD-710)
NDA NUMBER: NDA 21-677 / N000
DRUG NAME: Alimta (Pemetrexed, LY231514)
INDICATION: Locally advanced or metastatic non-small cell lung cancer
SPONSOR: Eli Lilly and Company

3. Based on our internal discussion, we have determined that the lower limit of the 95% confidence interval is probably too conservative for estimating the control treatment (docetaxel) effect; thus, the resulting statistical test used in the statistical review dated 6/29/04 is also probably too conservative.

Design and analysis of non-inferiority mortality trials in oncology

Mark Rothmann[1,*,†], Ning Li[1], Gang Chen[1], George Y. H. Chi[1], Robert Temple[2] and Hsiao-Hui Tsou[1]

Statist. Med. 2003; **22**:239–264 (DOI: 10.1002/sim.1400)

Good Enough: A Primer on the Analysis and Interpretation of Noninferiority Trials

Sanjay Kaul, MD, and George A. Diamond, MD

Ann Intern Med. 2006;145:62-69

Active-control noninferiority trials are being performed with increasing frequency, especially in cardiovascular and oncologic applications when placebo-controlled trials are considered unethical.

Because of this uncertainty, the noninferiority margin is typically defined in terms of some fraction (f) of the standard treatment effect to be preserved (9, 11).

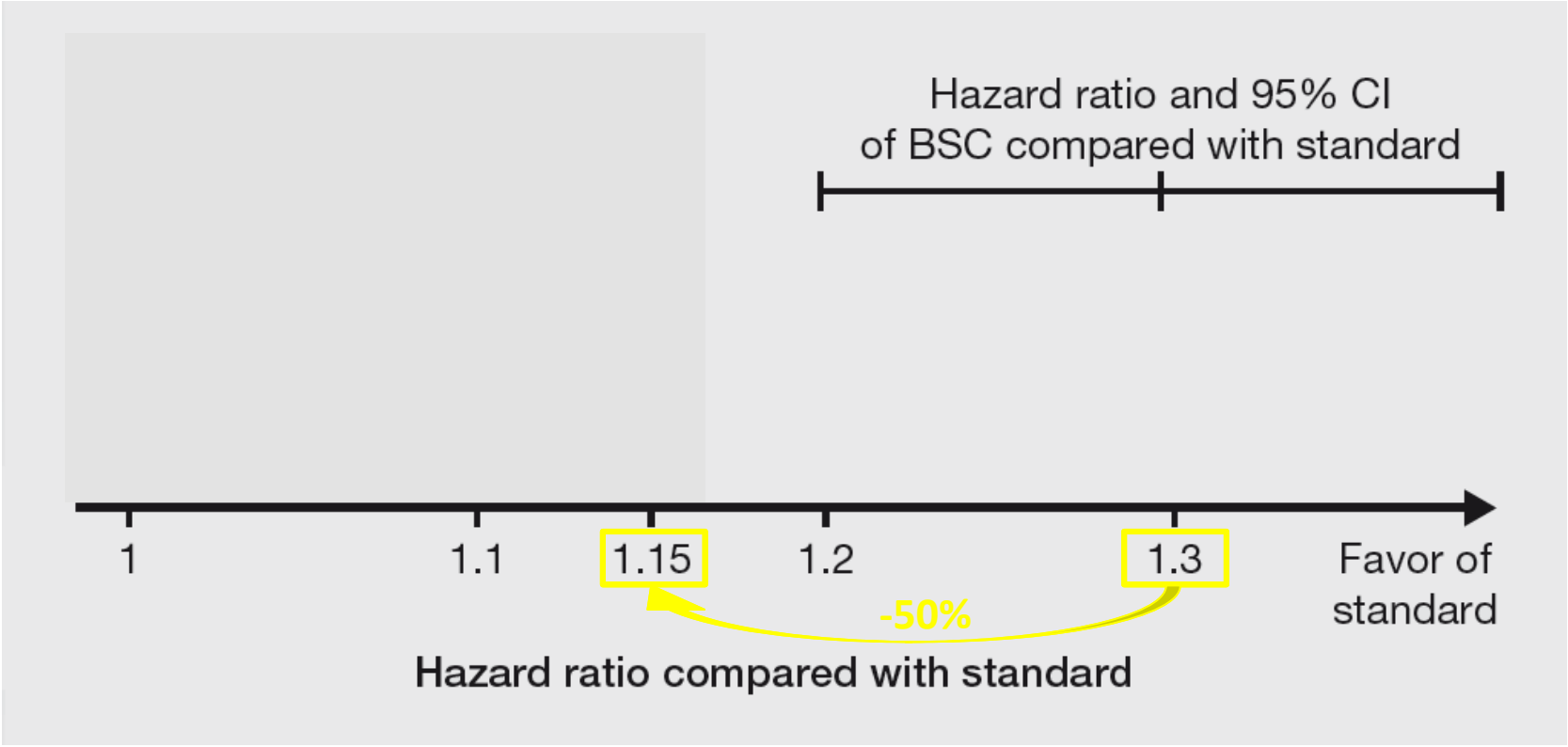
In the context of oncologic and thrombolytic trials, when mortality is evaluated, the U.S. Food and Drug Administration has suggested an f value of 0.5 (9, 11, 16).

Statistical Issues and Recommendations for Noninferiority Trials in Oncology: A Systematic Review

Shiro Tanaka[1], Yousuke Kinjo[2], Yoshiki Kataoka[2], Kenichi Yoshimura[1], and Satoshi Teramukai[1]

Clin Cancer Res; 18(7); 1837–47. ©2012 AACR.

Percent Retention Method

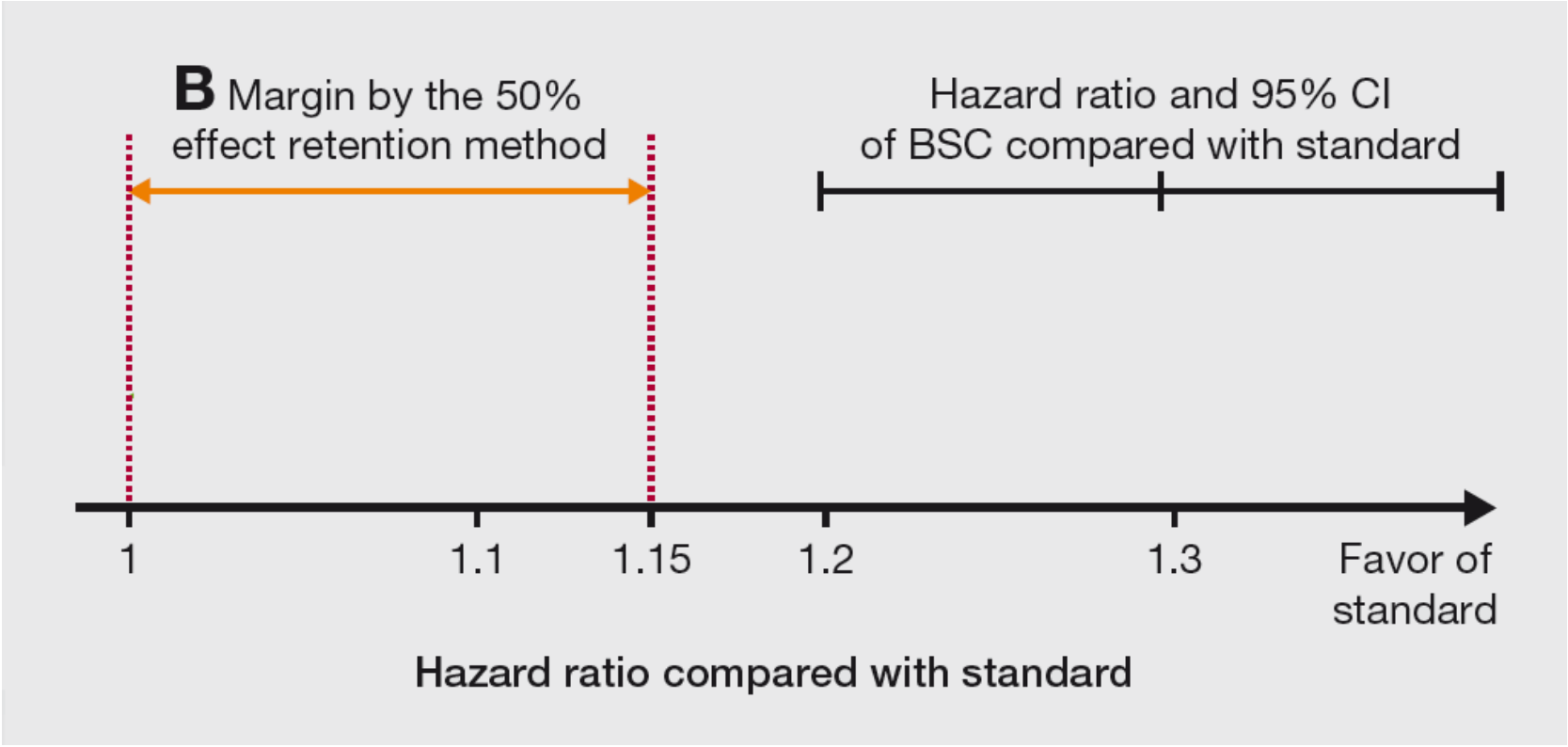


Statistical Issues and Recommendations for Noninferiority Trials in Oncology: A Systematic Review

Shiro Tanaka[1], Yousuke Kinjo[2], Yoshiki Kataoka[2], Kenichi Yoshimura[1], and Satoshi Teramukai[1]

Clin Cancer Res; 18(7); 1837–47. ©2012 AACR.

Percent Retention Method



Highlights from the Tenth World Conference on Lung Cancer

TRACEY L. EVANS

The University of Pennsylvania Medical Center—Presbyterian, Division of Hematology/Oncology, Philadelphia, Pennsylvania, USA

The Oncologist 2004;9:232-238

	Pemetrexed (n = 283)	Docetaxel (n = 288)
Median survival	8.3 months	7.9 months
Hazard ratio* (95% CI)	0.99 (0.8-1.2)	
1-year survival rate	29.7%	29.7%

*This study was initially designed to be a noninferiority trial using the fixed margin method. Using this statistical test, pemetrexed would be declared no worse than 10% less efficacious than docetaxel if the upper limit of the 95% CI of the overall survival hazard ratio was ≤1.11. The upper limit of the hazard ratio observed in the study was 1.2. By another retrospectively applied test of noninferiority, the percent retention method, the comparator must demonstrate preservation of no less than 50% of the benefit of the standard arm. Pemetrexed in this study had a 102% retention of efficacy (95% CI = 52%-157%, p value = 0.047). Therefore, by this test, pemetrexed would be declared noninferior.

Randomized Phase III Trial of Pemetrexed Versus Docetaxel in Patients With Non–Small-Cell Lung Cancer Previously Treated With Chemotherapy

Nasser Hanna, Frances A. Shepherd, Frank V. Fossella, Jose R. Pereira, Filippo De Marinis, Joachim von Pawel, Ulrich Gatzemeier, Thomas Chang Yao Tsao, Miklos Pless, Thomas Muller, Hong-Liang Lim, Christopher Desch, Klara Szondy, Radj Gervais, Shaharyar, Christian Manegold, Sofia Paul, Paolo Paoletti, Lawrence Einhorn, and Paul A. Bunn Jr.

J Clin Oncol 22:1589-1597. © 2004 by American Society of Clinical Oncology

the hypothesis that pemetrexed retained ≥ 50% of the survival benefit of docetaxel over BSC using data from the randomized comparative trial of docetaxel versus BSC by Shepherd et al[5] was prospectively planned (percent retention method).[14]

Apixaban versus Warfarin in Patients with Atrial Fibrillation

Christopher B. Granger, M.D., John H. Alexander, M.D., M.H.S., John J.V. McMurray, M.D., Renato D. Lopes, M.D., Ph.D., Elaine M. Hylek, M.D., M.P.H., Michael Hanna, M.D., Hussein R. Al-Khalidi, Ph.D., Jack Ansell, M.D., Dan Atar, M.D., Alvaro Avezum, M.D., Ph.D., M. Cecilia Bahit, M.D., Rafael Diaz, M.D., J. Donald Easton, M.D., Justin A. Ezekowitz, M.B., B.Ch., Greg Flaker, M.D., David Garcia, M.D., Margarida Geraldes, Ph.D., Bernard J. Gersh, M.D., Sergey Golitsyn, M.D., Ph.D., Shinya Goto, M.D., Antonio G. Hermosillo, M.D., Stefan H. Hohnloser, M.D., John Horowitz, M.D., Puneet Mohan, M.D., Ph.D., Petr Jansky, M.D., Basil S. Lewis, M.D., Jose Luis Lopez-Sendon, M.D., Prem Pais, M.D., Alexander Parkhomenko, M.D., Freek W.A. Verheugt, M.D., Ph.D., Jun Zhu, M.D., and Lars Wallentin, M.D., Ph.D., for the ARISTOTLE Committees and Investigators*

10.1056/NEJMoa1107039 NEJM.ORG

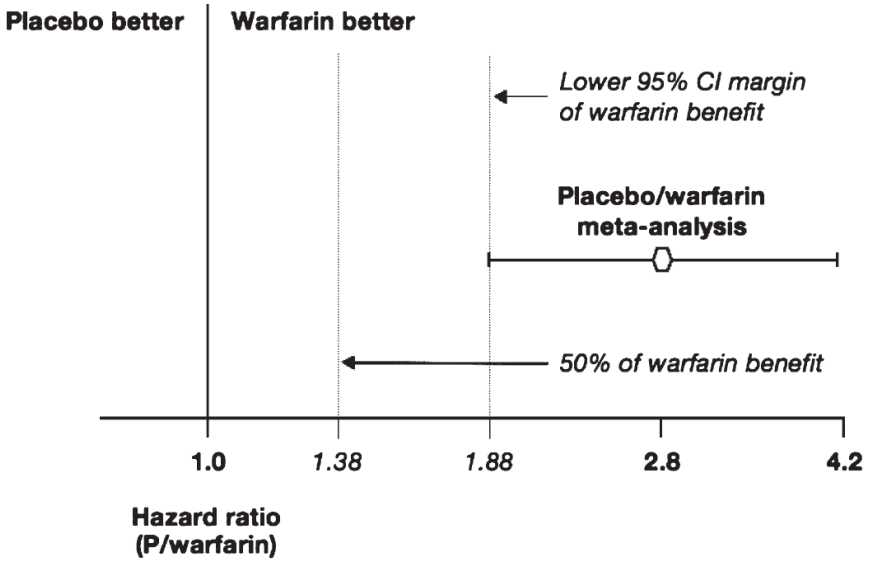
The primary noninferiority hypothesis required that [redacted] (62%) in six previous, major randomized, controlled trials.[10] This hypothesis provided a lower 95% confidence interval of 1.88 for the relative risk with placebo as compared with warfarin, and one half of this value was [redacted]

Antithrombotic drug development for atrial fibrillation: Proceedings, Washington, DC, July 25-27, 2005

Kevin Jackson, MD,[a] Bernard J. Gersh, MB, ChB, DPhil,[b] Norman Stockbridge, MD, PhD,[c] Thomas R. Fleming, PhD,[d] Robert Temple, MD,[c] Robert M. Califf, MD,[a] Stuart J. Connolly, MD,[e] Lars Wallentin, MD, PhD,[f] and Christopher B. Granger, MD[a] Participants in the Duke Clinical Research Institute/ American Heart Journal Expert Meeting on Antithrombotic Drug Development for Atrial Fibrillation

Durham, NC; Rochester, MN; Silver Spring, MD; Seattle, WA; Hamilton, Ontario, Canada; and Uppsala, Sweden

Am Heart J 2008;155:829-40



The RRR of warfarin compared with placebo in these trials using a random effects model was 0.36 (95% CI 0.24-0.53), such that the inverse of the upper boundary (ie, control compared with warfarin) is 1.88 (1/0.53). To establish that at least half of the warfarin effect is preserved, the noninferiority margin is 1.88 or 1.38 (ie, the margin is the midpoint between 1.0 and 1.88 on a log scale rather than linear scale because the primary parameter estimated is the logarithm of the relative risk).

Phase III Non-Inferiority Study of Cabazitaxel 20 mg/m^2 versus Cabazitaxel 25 mg/m^2 in Patients with Metastatic Castration-Resistant Prostate Cancer Previously Treated with Docetaxel (PROSELICA)

Johann de Bono,[1] Anne-Claire Hardy-Bessard,[2] Choung Soo Kim,[3] Lajos Geczi,[4] Daniel Ford,[5] Loic Mourey,[6] Joan Carles,[7] Phillip Parente,[8] Albert Font,[9] Gabriel Kacso,[10] Mustapha Chadjaa,[11] Wenping Zhang,[12] John Bernard,[13] Mario Eisenberger[14]

[1]Royal Marsden NHS Foundation Trust and The Institute of Cancer Research, Sutton, UK; [2]Centre Armoricain d'Oncologie, CARIO, Plérin, France; [3]Asan Medical Center, Seoul, South Korea; [4]National Institute of Oncology, Budapest, Hungary; [5]City Hospital, Cancer Centre Queen Elizabeth Hospital, Birmingham, UK; [6]Institut Claudius Regaud, IUCT-O, Toulouse, France; [7]Vall d'Hebron University Hospital, Vall d'Hebron Institute of Oncology, Barcelona, Spain; [8]Eastern Health Clinical School, Monash University, Box Hill Hospital, Australia; [9]Institut Català d'Oncologia, Hospital Universitari Germans Trias i Pujol, Badalona, Spain; [10]Amethyst Radiotherapy Center, Cluj, Romania; [11]Sanofi, Vitry-sur-Seine, France; [12]Sanofi Genzyme, Bridgewater, NJ, USA; [13]Sanofi Genzyme, Cambridge, MA, USA; [14]The Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins, Baltimore, MD, USA

PRESENTED AT: ASCO ANNUAL MEETING '16

Presented by: Johann de Bono

Presented By Johann De Bono at 2016 ASCO Annual Meeting

PROSELICA: Non-inferiority Design

- To test if the 95% OS HR CI of CBZ 20 and CBZ 25 are within the predefined parameters requested by the FDA
- The basic hypothesis is that CBZ 20 “maintains at least 50% of the OS benefit of CBZ 25 relative to mitoxantrone observed on the TROPIC study”
- To claim non-inferiority with 95% confidence level in the final analysis, the upper-bound CI of HR for CBZ 20 versus CBZ 25 could not exceed 1.214 under one-sided 98.89% confidence level after being adjusted for the interim analyses

Some essential considerations in the design and conduct of non-inferiority trials

Thomas R Fleming[a,b], *Katherine Odem-Davis*[a,b], *Mark D Rothmann*[c] *and Yuan Li Shen*[c]

Clinical Trials 2011; **8**: 432–439

Irregularities in quality of the conduct of the non-inferiority trial induce increased risk of both bias and variability.

While such irregularities are of concern in superiority trials, they are even more problematic in a non-inferiority trial since they often dilute the sensitivity to true differences between the experimental intervention and Standard regimens, leading to an increased risk of falsely declaring non-inferiority in settings where the test treatment truly is clinically inferior to Standard.

Studi di non-inferiorità:
Analisi ITT *Vs* analisi PP

The European Agency for the Evaluation of Medicinal Products
Evaluation of Medicines for Human Use
London, 27 July 2000 CPMP/EWP/482/99

IV.2.3 Choice of analysis set

In a superiority trial the full analysis set, based on the ITT (intention-to-treat) principle, is the analysis set of choice, with appropriate support provided by the PP (per protocol) analysis set.

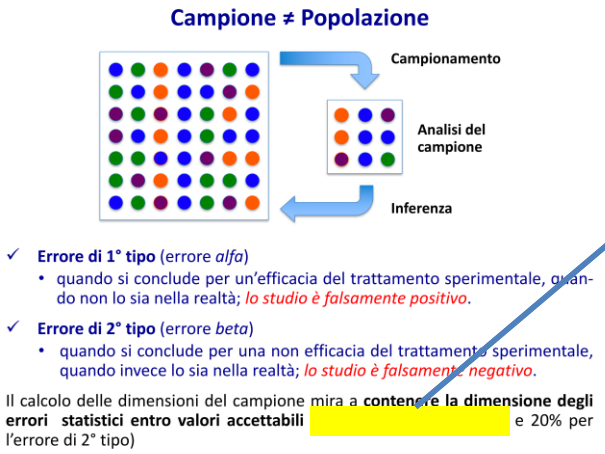
Guidance for Industry
Non-Inferiority Clinical Trials

March 2010
Clinical/Medical

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)

Although an "as-treated" analysis is therefore often suggested as the primary analysis for NI studies, there are also significant concerns with the possibility of informative censoring in an as-treated analysis.

Multiplicity



Probability of at least one false significant result

Number of tests	Probability
2	0.098
5	0.226
10	0.401
50	0.923

Multiple comparisons, **multiplicity** or multiple testing problem occurs when one **considers a set of statistical inferences simultaneously or infers a subset of parameters selected based on the observed values**

modif from Schulz & Grimes; Lancet 2005; 365: 1657–61

Apixaban for Extended Treatment of Venous Thromboembolism

Giancarlo Agnelli, M.D., Harry R. Buller, M.D., Ph.D., Alexander Cohen, M.D., Madelyn Curto, D.V.M., Alexander S. Gallus, M.D., Margot Johnson, M.D., Anthony Porcari, Ph.D., Pharm.D., Gary E. Raskob, Ph.D., and Jeffrey I. Weitz, M.D., for the AMPLIFY-EXT Investigators*

N Engl J Med 2012.

STATISTICAL ANALYSIS

The study was designed to test the hypothesis that

with respect to the primary efficacy outcome. Assuming an estimated incidence in the placebo group of 6.8% at 12 months and a decrease in the

A sharper Bonferroni procedure for multiple tests of significance

BY YOSEF HOCHBERG

Biometrika (1988), **75**, 4, *pp.* 800–2

for the study to have 90% powe e su-
periority of apixaban over placebo, at a tw sided
alpha level of 0.05,

20

Multiplicity in randomised trials II: subgroup and interim analyses

Kenneth F Schulz, David A Grimes

Lancet 2005; 365: 1657–61

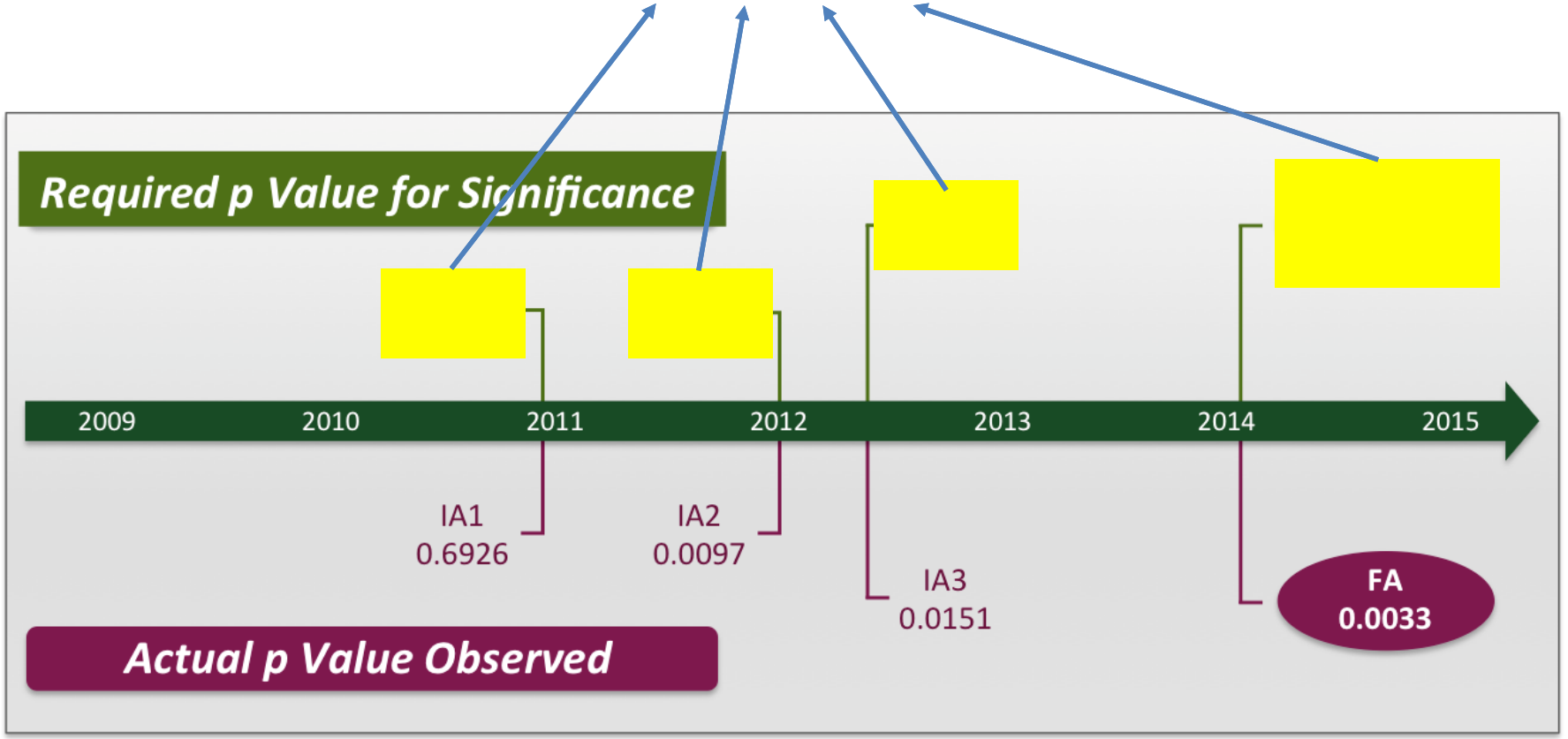
Number of planned interim analyses	Interim analysis	Pocock	Peto	O'Brien-Fleming
2	1	0·029	0·001	0·005
	2 (final)	0·029	0·05	0·048
3	1	0·022	0·001	0·0005
	2	0·022	0·001	0·014
	3 (final)	0·022	0·05	0·045
4	1	0·018	0·001	0·0001
	2	0·018	0·001	0·004
	3	0·018	0·001	0·019
	4 (final)	0·018	0·05	0·043
5	1	0·016	0·001	0·00001
	2	0·016	0·001	0·0013
	3	0·016	0·001	0·008
	4	0·016	0·001	0·023
	5 (final)	0·016	0·05	0·041

Overall α=0·05.

Table 2: Interim stopping levels (p values) for different numbers of planned interim analyses by group sequential design[14,15]

Il piano di analisi dello studio COU-302…

Livello di significatività statistica per rPFS: 0.01
Dimensione residua dell'errore di 1° tipo per OS <0.04



The European Agency for the Evaluation of Medicinal Products
Evaluation of Medicines for Human Use

London, 19 September 2002
CPMP/EWP/908/99

COMMITTEE FOR PROPRIETARY MEDICINAL PRODUCTS (CPMP)

POINTS TO CONSIDER ON MULTIPLICITY ISSUES IN CLINICAL TRIALS

2. ADJUSTMENT FOR MULTIPLICITY – WHEN IS IT NECESSARY AND WHEN IS IT NOT?

Sometimes a series of related objectives is pursued in the same trial, where one objective is of greatest importance but convincing results in others would clearly add to the value of the treatment.

In these situations, and, consequently, level α, i.e. no reduction is necessary.

(and confidence intervals may be provided) The hierarchical order may be a natural one (e.g. hypotheses are ordered in time or with respect to the seriousness of the considered variables) or may result from the particular interests of the investigator. Again, no reduction or splitting of α is necessary.

Effect of a monoclonal antibody to PCSK9, REGN727/SAR236553, to reduce low-density lipoprotein cholesterol in patients with heterozygous familial hypercholesterolaemia on stable statin dose with or without ezetimibe therapy: a phase 2 randomised controlled trial

Evan A Stein, Dan Gipe, Jean Bergeron, Daniel Gaudet, Robert Weiss, Robert Dufour, Richard Wu, Robert Pordy

Lancet 2012; 380: 29–36

To address the multiple comparisons of the four treatment groups compared with placebo for the primary efficacy endpoint analysis, we applied a hierarchical testing procedure to ensure strong control of the overall type-1 error rate at the two-sided 0·05 significance level.

Effect of a monoclonal antibody to PCSK9, REGN727/SAR236553, to reduce low-density lipoprotein cholesterol in patients with heterozygous familial hypercholesterolaemia on stable statin dose with or without ezetimibe therapy: a phase 2 randomised controlled trial

Evan A Stein, Dan Gipe, Jean Bergeron, Daniel Gaudet, Robert Weiss, Robert Dufour, Richard Wu, Robert Pordy
Lancet 2012; 380: 29–36

To address the multiple comparisons of the four treatment groups compared with placebo for the primary efficacy endpoint analysis, we applied a hierarchical testing procedure to ensure strong control of the overall type-1 error rate at the two-sided 0·05 significance level.

The order used was REGN727 150 mg every 2 weeks versus placebo first; REGN727 300 mg every 4 weeks versus placebo second; REGN727 200 mg every 4 weeks versus placebo third; and finally, REGN727 150 mg every 4 weeks versus placebo.

The hierarchical testing sequence continued only when the higher-order test was statistically significant at the two-sided 5% significant level.

Monotherapy with the PCSK9 inhibitor alirocumab versus ezetimibe in patients with hypercholesterolemia: Results of a 24 week, double-blind, randomized Phase 3 trial

Eli M. Roth [a,*], Marja-Riitta Taskinen [b], Henry N. Ginsberg [c], John J.P. Kastelein [d], Helen M. Colhoun [e], Jennifer G. Robinson [f], Laurence Merlet [g], Robert Pordy [h], Marie T. Baccara-Dinet [i]

International Journal of Cardiology 176 (2014) 55–61

Percent change from baseline in secondary lipid parameters (ITT and on-treatment analysis).

LS mean (SE) % change from baseline to week 24	Alirocumab versus ezetimibe		
	LS mean difference (SE) %	95% CI	*p*-Value
ITT			
Lp(a)[b]	−4.4 (5.3)	−14.8 to 5.9	0.4013
TGs[b]	−1.2 (5.9)	−12.7 to 10.3	0.8433
HDL-C	4.4 (2.7)	−1.0 to 9.8	0.1116
Apo A-1	5.3 (2.2)	0.9 to 9.8	0.0196

Efficacy and Safety of Alirocumab in Reducing Lipids and Cardiovascular Events

Jennifer G. Robinson, M.D., M.P.H., Michel Farnier, M.D., Ph.D., Michel Krempf, M.D., Jean Bergeron, M.D., Gérald Luc, M.D., Maurizio Averna, M.D., Erik S. Stroes, M.D., Ph.D., Gisle Langslet, M.D., Frederick J. Raal, M.D., Ph.D., Mahfouz El Shahawy, M.D., Michael J. Koren, M.D., Norman E. Lepor, M.D., Christelle Lorenzato, M.Sc., Robert Pordy, M.D., Umesh Chaudhari, M.D., and John J.P. Kastelein, M.D., Ph.D., for the ODYSSEY LONG TERM Investigators*

N Engl J Med 2015;372:1489-99.

Table 2. Primary Efficacy End Point and Selected Secondary Efficacy End Points in the Intention-to-Treat Population.*

End Point	Alirocumab (N=1530)	Placebo (N=780)	Alirocumab vs. Placebo		
			Least-Squares Mean Difference	95% CI	P Value
Primary efficacy end point: calculated LDL cholesterol					
Percentage change from baseline to wk 24	−61.0±0.7	0.8±1.0	−61.9±1.3	−64.3 to −59.4	
Selected secondary efficacy end points					
Proportion of patients who reached prespecified calculated LDL cholesterol levels by wk 24 — %‡					
<70 mg/dl in patients at very high risk or <100 mg/dl in patients at high risk	80.7	8.5			
<70 mg/dl regardless of risk	79.3	8.0			<0.001§
Apolipoprotein B	−52.8±0.7	1.2±1.0	−54.0±1.2	−56.3 to −51.7	
Total cholesterol	−37.8±0.5	−0.3±0.7	−37.5±0.8	−39.1 to −35.9	
Lipoprotein(a)¶	−29.3±0.7	−3.7±1.0	−25.6±1.3	−28.1 to −23.1	
Fasting triglycerides¶	−15.6±0.8	1.8±1.2	−17.3±1.4	−20.1 to −14.6	
HDL cholesterol	4.0±0.4	−0.6±0.5	4.6±0.7	3.3 to 5.9	
Apolipoprotein A1	4.0±0.4	1.2±0.6	2.9±0.7	1.6 to 4.2	

The P values are significant according to the fixed hierarchical approach used to ensure control of the overall type I error rate at the 0.05 level.

Nivolumab versus Everolimus in Advanced Renal-Cell Carcinoma

R.J. Motzer, B. Escudier, D.F. McDermott, S. George, H.J. Hammers, S. Srinivas, S.S. Tykodi, J.A. Sosman, G. Procopio, E.R. Plimack, D. Castellano, T.K. Choueiri, H. Gurney, F. Donskov, P. Bono, J. Wagstaff, T.C. Gauler, T. Ueda, Y. Tomita, F.A. Schutz, C. Kollmannsberger, J. Larkin, A. Ravaud, J.S. Simon, L.-A. Xu, I.M. Waxman, and P. Sharma, for the CheckMate 025 Investigators*

N Engl J Med 2015;373:1803-13

Survival was compared between the treatment groups with the use of the monitoring feature of East software, version 5.4 (Cytel), which is based on the Lan–DeMets with an O'Brien–Fleming stopping boundary used to reject the null hypothesis (i.e., that there is no treatment difference), while maintaining a two-sided overall alpha level of 0.05. Interim overall survival was projected at a 0.0148 nominal significance level; if the results for overall survival were significant at that level, the study could be stopped at the recommendation of the data monitoring committee and declared to be positive for efficacy.

If superiority with regard to the primary end point was demonstrated, a (estimated along with the exact 95% confidence interval with the use of the Clopper–Pearson method) at an alpha level of 0.05.

Statistical Analysis Plan

(to maintain a 2-sided alpha level of 0.05)

1. Overall Survival (interim)
2. Overall Survival (final)

alpha spending

3. Objective Response Rate
4. Progression Free Survival

hierarchical procedure