

Con il Patrocinio di



STUDI CLINICI: METODOLOGIA

Coordinatore:
Dr.ssa Stefania Gori

Evento ECM MODULO 1
(formazione di base)

"A good foundation"



NEGRAR
7-8 Settembre 2016

Centro Formazione
Ospedale Sacro Cuore
Don Calabria

Giovedì 8 Settembre 2016

- **Principi di dimensionamento campionario**

- **Gli errori statistici**

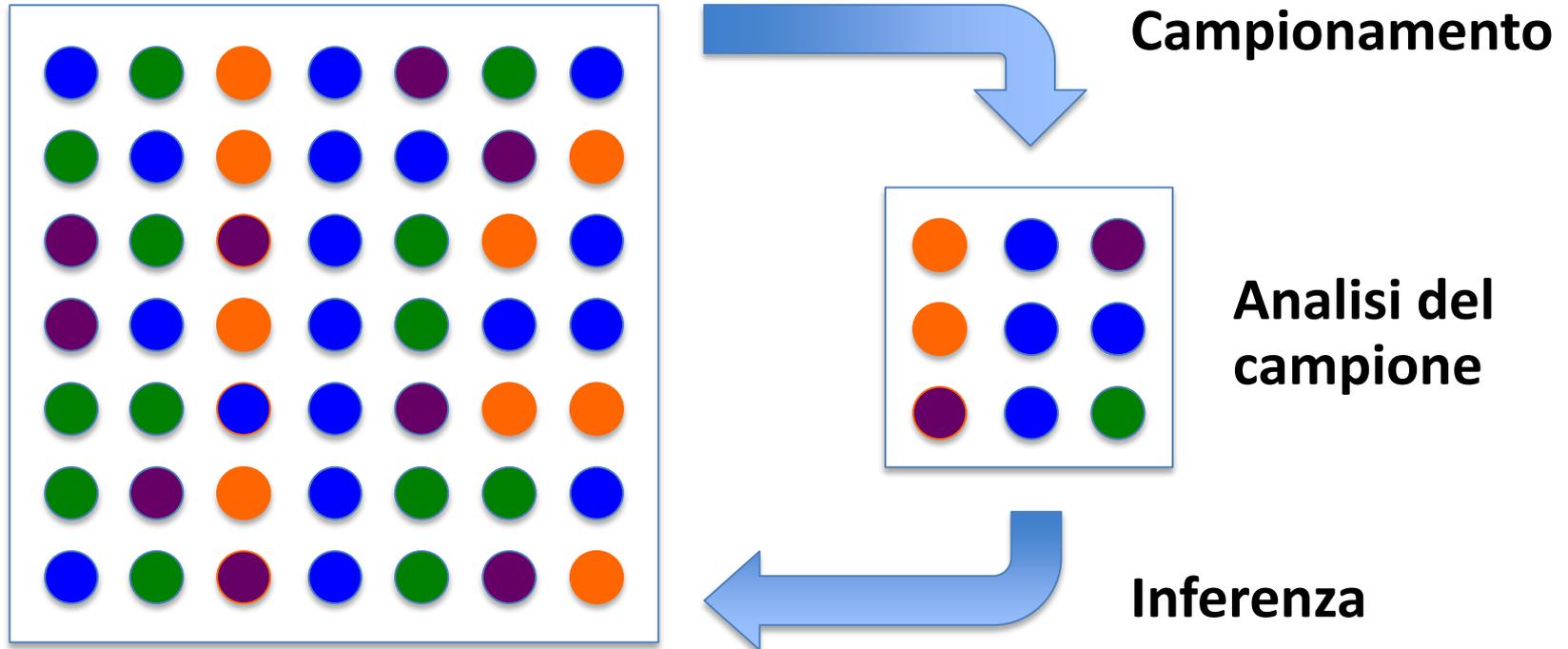
- ✓ *alpha error* (tipo I)
- ✓ *beta error* (tipo II)

- Il target di rilevanza clinica

- ✓ *delta* (di superiorità)
- ✓ *margin* *M* (di non inferiorità)

- Calcolo del campione i diversi tipi di variabili statistiche

Campione \neq Popolazione



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

POPOLAZIONE, CAMPIONE, INFERENZA

- ✓ La **Popolazione** comprende *tutti* i soggetti con le caratteristiche di interesse per un determinato studio
- ✓ Il **Campione** su cui lo studio verrà condotto potrà essere (o non essere) rappresentativo di tutte le caratteristiche della popolazione (la cosa ci è *ignota*)
- ✓ Lo Sperimentatore è naturalmente portato a trasferire alla *Popolazione* quanto osservato nel *Campione* (**Inferenza**)
- ✓ Il risultato del test statistico di confronto fra due trattamenti è influenzato (***ma non sappiamo a favore di chi, e in che misura***) dal fatto che lo studio è condotto su un campione

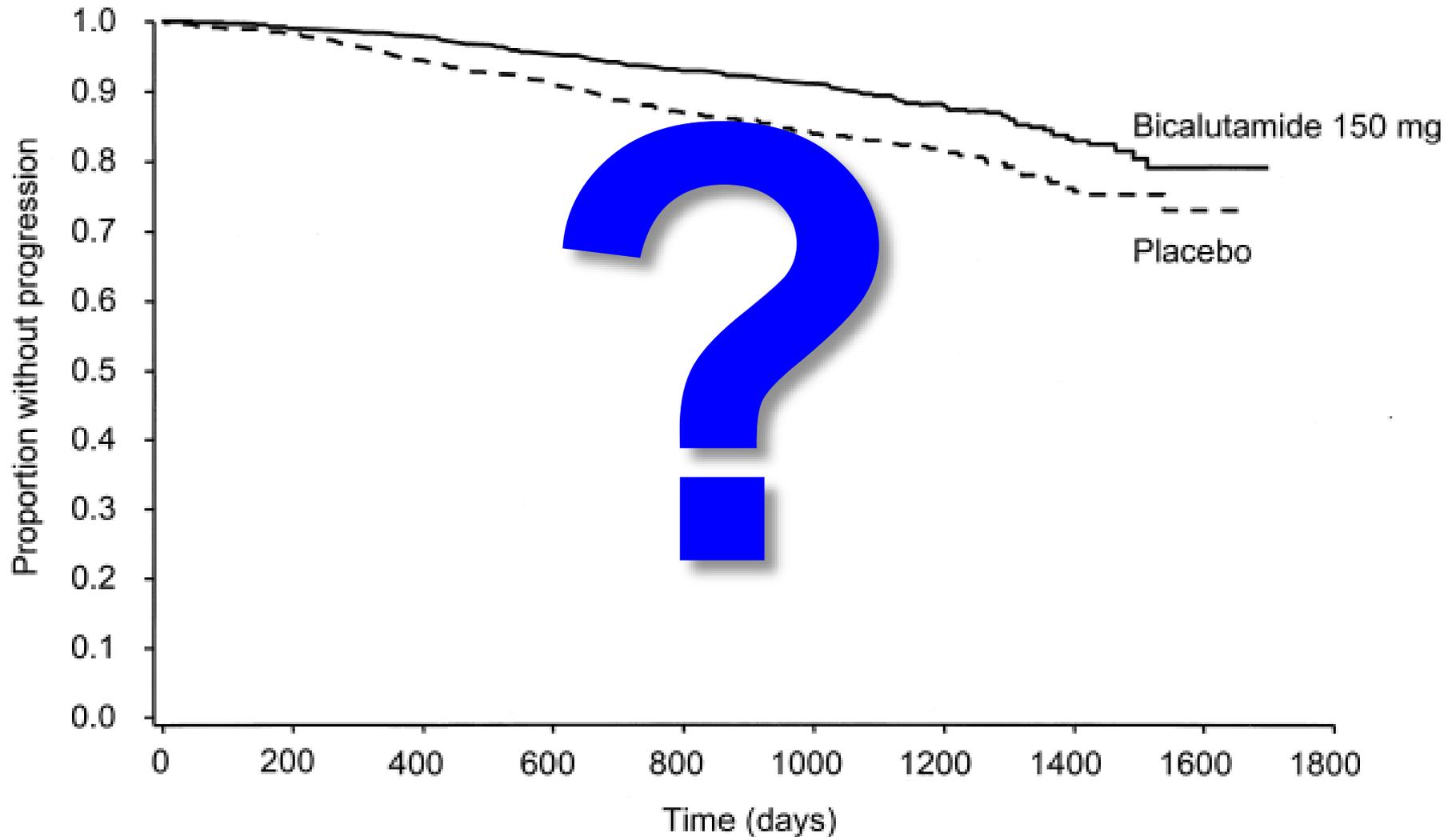
VARIABILE DI TIPO QUALITATIVO...

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 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.**

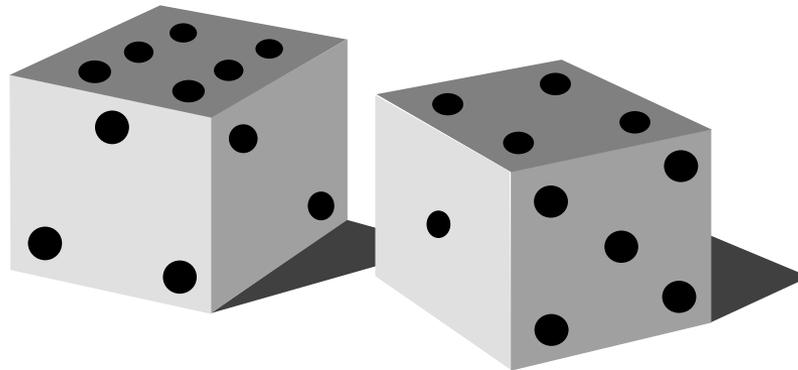
***effetto osservato = effetto del
trattamento + effetto del caso***

(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 (i.e. quanto 'pesa' il 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)

of Hypothesis

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

	H_0 is True	H_0 is NOT True
Accept H_0		
Reject H_0		

Il test statistico può risultare positivo o negativo

N.B.: Ipotesi di non differenza tra i trattamenti (H_0)

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	
Condannare il colpevole	

Making Decisions: Test of Hypothesis

	H_0 is True	H_0 is NOT True
Accept H_0	☺	
Reject H_0		☺

Il test potrebbe quindi dare
la risposta corretta...

**N.B.: Ipotesi di non differenza
tra i trattamenti (H_0)**

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_1
Condannare il colpevole	Corretto rifiuto di H_0
Condannare l'innocente	
Assolvere il colpevole	

Making Decisions: Test of Hypothesis

	H_0 is True	H_0 is NOT True
Accept H_0		☹️
Reject H_0	☹️	

... oppure potrebbero essere stati commessi degli errori.

N.B.: Ipotesi di non differenza tra i trattamenti (H_0)

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_1
Condannare il colpevole	Corretto rifiuto di H_0
Condannare l'innocente	Errore di 1° tipo
Assolvere il colpevole	Errore di 2° tipo

Making Decisions: Test of Hypothesis

	H_0 is True	H_0 is NOT True
Accept H_0		 → Type II error
Reject H_0	 ↓ Type I error	

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

N.B.: Ipotesi di non differenza tra i trattamenti (H_0)

ERRORI STATISTICI

- ✓ **Errore di 1° tipo (errore α : 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 β : 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)**

Con il Patrocinio di



STUDI CLINICI: METODOLOGIA

Coordinatore:
Dr.ssa Stefania Gori

Evento ECM MODULO 1
(formazione di base)

"A good foundation"



NEGRAR
7-8 Settembre 2016

Centro Formazione
Ospedale Sacro Cuore
Don Calabria

Giovedì 8 Settembre 2016

- **Principi di dimensionamento campionario**

- Gli errori statistici
 - ✓ *alpha error (tipo I)*
 - ✓ *beta error (tipo II)*
- Il target di rilevanza clinica
 - ✓ *delta (di superiorità)*
 - ✓ *marginale M (di non inferiorità)*
- **Calcolo del campione i diversi tipi di variabili statistiche**

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%

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 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 **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

Sample size and power calculations were based on the ability to detect a treatment difference in LDL-C of **40%** 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 **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

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 **20%** 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%.

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.

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

Matti J. Tikkanen, MD, PhD

Christina Lindahl, MD

Ole Faergeman, MD, DMSc

Ingar Holme, PhD

Michael Szarek, MS

John J. P. Kastelein, MD, PhD

Mogens Lytken Larsen, MD, DMSc

John Tsai, MD

Anders G. Olsson, MD, PhD

Fredrik S. Bendiksen, 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.

When performing 20 independent tests, we shall expect to have one significant result even though no difference exists.

$$1/20 = 0.05$$

Probability of at least one significant result

Number of tests	Probability
1	0.050
2	0.098
5	0.226
10	0.401
50	0.923

Nel caso di >1 test statistico sullo stesso campione, viene superata la probabilità (convenzionalmente accettata) che il risultato osservato sia falsamente positivo

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 each dose of apixaban would be superior to placebo 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

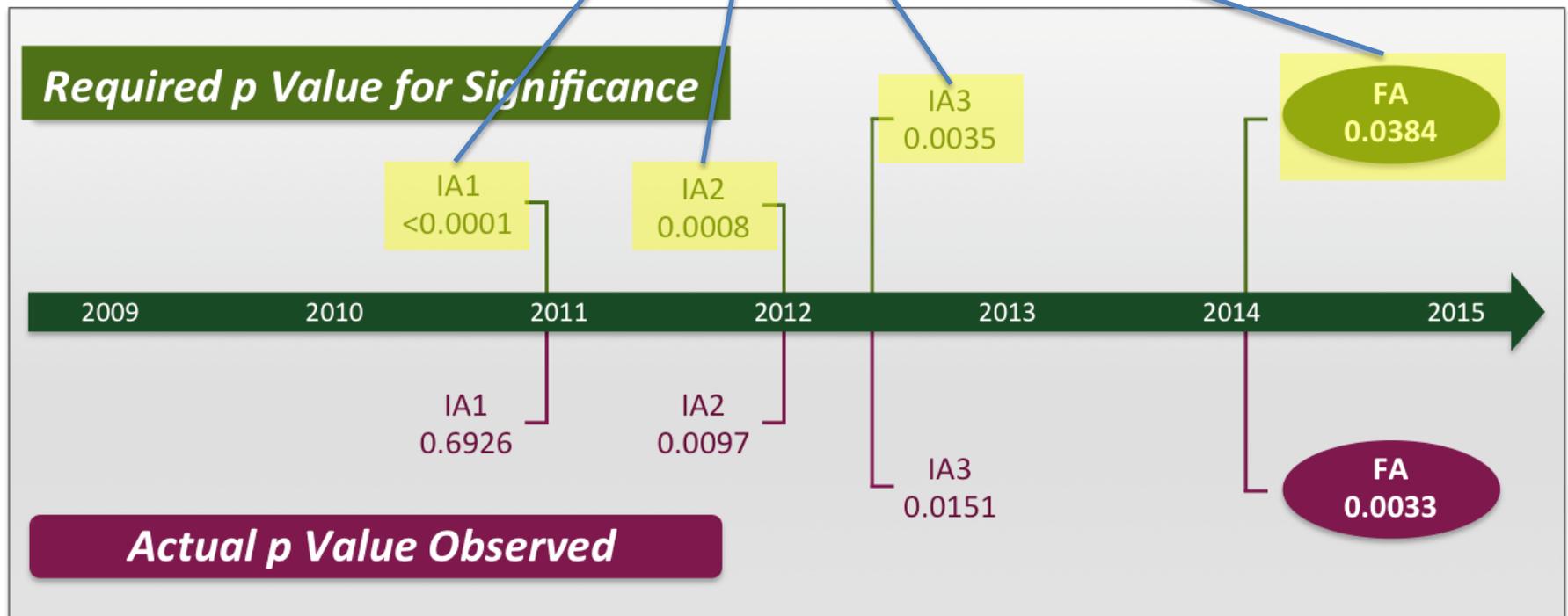
BY YOSEF HOCHBERG

Biometrika (1988), 75, 4, pp. 800-2

for the study to have 90% power to detect the superiority of apixaban over placebo, at a two-sided alpha level of 0.05, with the use of the Hochberg multiple-testing method.²⁰

Il piano di analisi dello studio COU-302...

Dimensione globale dell'errore di 1° tipo <0.05



**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, there is no intention or opportunity to select the most favourable result and, consequently, the individual type I error levels are set equal to the overall type I error level α , i.e. no reduction is necessary.

In such cases the hypotheses may be tested (and confidence intervals may be provided) according to a hierarchical strategy. 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. The hierarchical order for testing null hypotheses, however, has to be pre-specified in the study protocol.

**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			
Apo B	−25.8 (3.3)	−32.3 to −19.2	<0.0001 ^a
Non-HDL-C	−25.5 (4.1)	−33.5 to −17.4	<0.0001 ^a
Total cholesterol	−18.7 (3.0)	−24.7 to −12.7	<0.0001 ^a
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

^a Statistically significant according to the fixed hierarchical approach used to control overall type-I error rate.