

# Note statistiche. Sperimentazioni cliniche controllate randomizzate sia di superiorità che di non inferiorità: considerazioni critiche

Ettore Marubini, Paola Rebora, Giuseppe Reina

Istituto di Statistica Medica e Biometria, Università degli Studi, Milano

**Key words:**

Analysis per protocol;  
Intention-to-treat;  
Non-inferiority  
randomized controlled  
clinical trials;  
Superiority randomized  
controlled clinical trials.

The aim of this statistical note, the fourth in the series, is to describe and critically appraise the randomized controlled clinical trial (RCCT) GUSTO V, which combined in a single RCCT both the superiority and non-inferiority hypotheses. In this note we present the logical path that the authors have presumably followed in planning a RCCT of such a kind. The results are reported and possible critical aspects are highlighted and debated upon. Finally the cardiologist reader is stimulated to give his own opinion on this kind of approach.

(Ital Heart J Suppl 2005; 6 (6): 361-364)

© 2005 CEPI Srl

Questa nota è stata redatta nell'ambito di una ricerca supportata dal Consiglio Nazionale delle Ricerche (contratto n. CU02.00264.ST97).

Ricevuto il 31 gennaio 2005; nuova stesura il 12 aprile 2005; accettato il 13 aprile 2005.

**Per la corrispondenza:**

Prof. Ettore Marubini

Istituto di Statistica  
Medica e Biometria  
Università degli Studi  
Via Venezian, 1  
20133 Milano  
E-mail:

ettore.marubini@unimi.it

## Introduzione

Nella prima di queste note statistiche si sono discusse le misure di effetto assolute e relative correntemente utilizzate nelle sperimentazioni cliniche controllate randomizzate (SCCR) di superiorità<sup>1</sup>. Nella seconda nota l'attenzione è stata rivolta alle SCCR di non inferiorità<sup>2</sup>. La presente nota si propone l'obiettivo di discutere criticamente lo studio GUSTO V (Global Use of Strategies to Open Occluded Coronary Arteries)<sup>3</sup> (i cui risultati sono stati pubblicati nel 2001) che costituisce un tentativo di combinare in una sola sperimentazione le due tipologie testé menzionate. Infatti questa sperimentazione è stata pianificata con lo scopo di saggiare "whether the combination of half-dose reteplase and abciximab would be superior, or not inferior, to reteplase alone for mortality at 30 days after enrolment".

## Sperimentazione clinica controllata randomizzata GUSTO V

Per quanto attiene alla tipologia di superiorità, nel paragrafo "Statistical analysis" gli autori dello studio GUSTO V specificano: "About 16 600 patients were required to detect a 15% relative mortality reduction from 7.4 to 6.3% ... with ... 80% power [... ad un livello di confidenza al

95%, aggiunta degli estensori di questa nota]".

Sembra ragionevole pensare che i risultati della SCCR GUSTO III<sup>4</sup>, pubblicati nel 1997, rappresentino il punto di partenza per la programmazione della SCCR GUSTO V. Nella SCCR GUSTO III è stato riferito che la mortalità nel gruppo trattato con reteplase (R) era pari a 7.47% (approssimato a 7.4% =  $\mu_R$ ). La diminuzione relativa di mortalità del 15% ( $\delta_R$ , come introdotta nella prima nota statistica), prevista per il trattamento con reteplase + abciximab (R + A) rispetto a quello con solo reteplase, rappresenta la quantità clinicamente rilevante a supporto dell'ipotesi di superiorità formulata dagli autori, precisamente  $\delta_R = -15\%$ . Il 15% di 7.4% è pari a 1.11% e ne consegue che la mortalità per il trattamento reteplase + abciximab dovrebbe essere uguale a:

$$7.4\% - 1.11\% = 6.29\% \cong 6.3\% = \mu_{R+A}$$

Nella prima nota<sup>1</sup> la *potenza* di uno studio è stata definita come un indicatore della capacità dello studio di porre in evidenza una differenza tra i due trattamenti, pari al valore clinicamente rilevante ( $\delta_R$ ) previsto nella programmazione, quando tale differenza *effettivamente esista* nella popolazione obiettivo che si considera sottoposta per metà all'uno e per metà all'altro dei due trattamenti. Questo indicatore assume valori che, a partire dal 2.5% (o 5%), possono tendere al 100%.

La SCCR GUSTO V è stata programmata con una dimensione dello studio di 8300 soggetti per braccio a mettere in evidenza una diminuzione relativa ( $\delta_R = -15\%$ ) con una potenza dell'80%.

Gli autori della SCCR GUSTO V continuano affermando: "... For the combination of half-dose reteplase and abciximab to be regarded as non-inferior, the upper bound of the 95% CI for the relative risk could be no greater than 1.10".

Dunque, lo studio di superiorità è stato programmato in termini di riduzione relativa di mortalità ( $\delta_R = -15\%$ ), mentre lo studio di non inferiorità in termini di rischio relativo ( $\rho_R = 1.10$ ). Per rendere uniforme l'interpretazione, esporremo le considerazioni che seguono in termini di rischio relativo per entrambe le tipologie di studio.

Per quanto attiene alla SCCR di non inferiorità, è opportuno ricordare che in tale tipo di studio il ricercatore desidera dimostrare che l'effetto del nuovo trattamento (R + A) non è clinicamente peggiore di quello del suo controllo attivo (in questo caso il solo R): tale effetto può essere superiore o anche di poco inferiore a quello del controllo attivo, purché l'inferiorità non superi una certa quantità clinicamente tollerabile posta uguale a 1.1, in termini di rischio relativo, nella SCCR GUSTO V.

Si tratta ora di illustrare il percorso logico atto a determinare il margine di tolleranza su riferito. Gli autori vogliono ancorare la scelta di tale margine alla dimensione dell'effetto del controllo attivo nei confronti del placebo (P) ricavabile dai dati storici. Poiché non esistono risultati di SCCR precedenti in cui l'effetto di reteplase sia stato valutato nei confronti del placebo, essi assumono che reteplase abbia un'efficacia pari a quella di streptochinasi (SK). Ciò è compatibile con:

- i risultati della già citata SCCR GUSTO III, dai quali si evince che la mortalità del gruppo trattato con reteplase è stata pari a circa il 7.4% ( $\mu_R \cong 7.4\%$ );
- il risultato della SCCR GUSTO<sup>5</sup>, che confrontò l'effetto di alteplase con quello della streptochinasi, dal quale è emerso che la mortalità nel gruppo dei trattati con streptochinasi è risultata uguale a  $\mu_{SK} = 7.4\%$ .

Ciò consente agli autori di GUSTO V di porre l'uguaglianza tra le seguenti differenze di effetti:

$$\text{placebo-reteplase} = \text{placebo-streptochinasi} \\ \text{ovvero:}$$

$$\mu_P - \mu_R = \mu_P - \mu_{SK}$$

La differenza di mortalità degli effetti di placebo-streptochinasi si ricava dai risultati dello studio ISIS-2 (International Study of Infarct Survival-2)<sup>6</sup> ed è  $\mu_P - \mu_{SK} = 2.8\%$ .

La somma di:

$$\mu_R + 2.8\% = 7.4\% + 2.8\% = 10.2\% = \mu_P$$

permette di ipotizzare quale valore avrebbe la mortalità del gruppo placebo, nel caso che il placebo fosse somministrato al terzo gruppo della SCCR GUSTO V (si veda nota statistica 2)<sup>2</sup>. Il margine in termini assoluti per la

"... preservation of at least 75% of the benefit of streptokinase over placebo ..." può ora essere calcolato come:

$$10.2\% - 0.75 \cdot 2.8\% = 8.1\%$$

Ciò significa che gli autori di GUSTO V sono disposti a tollerare una mortalità del trattamento con R + A ( $\mu_{R+A} = 8.1\%$ ), superiore a quella di R ( $\mu_R = 7.4\%$ ), ma non superiore al valore 8.1%. In termini di rischio relativo il margine diviene quindi:

$$\frac{\mu_{R+A}}{\mu_R} = \frac{8.1\%}{7.4\%} = 1.09 \cong 1.1$$

Dal momento che la SCCR GUSTO V era stata programmata sia come studio di superiorità che come studio di non inferiorità, la regola decisionale, in termini di rischio relativo, adatta a considerare contemporaneamente queste due tipologie di studio è illustrata nella figura 1. In entrambi i pannelli è evidenziata la regione di accettazione dell'ipotesi clinica: di superiorità (a) e di non inferiorità (b) (si veda la prima e la seconda di queste note statistiche<sup>1,2</sup>).

Dal paragrafo "Results" si ricava che i pazienti trattati con reteplase furono 8260 e quelli trattati con reteplase + abciximab 8328.

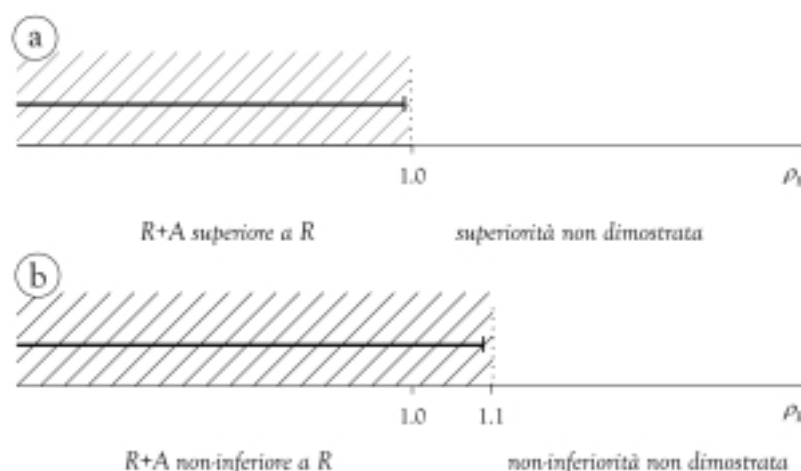
Inoltre si legge: "All-cause mortality at 30 days was 488 (5.9% [ $m_R$ ]) for the reteplase group, compared with 468 (5.6% [ $m_{R+A}$ ]) for the combined reteplase and abciximab group". Se ne deduce che il rischio relativo è risultato pari a:

$$\frac{m_{R+A}}{m_R} = \frac{\frac{468}{8328}}{\frac{488}{8260}} = \frac{5.6\%}{5.9\%} = 0.95$$

con intervallo di confidenza di  $\rho_R$  al 95%: [0.84; 1.08]. Poiché l'estremo superiore dell'intervallo di confidenza al 95% di  $\rho_R$ , 1.08, è  $> 1$  e  $< 1.1$ , gli autori non possono concludere a favore della non inferiorità del trattamento reteplase + abciximab rispetto al solo reteplase.

### Considerazioni finali

Nel documento CPMP/EWP/482/99<sup>7</sup> (pubblicato da EMEA-The European Agency for the Evaluation of Medicinal Products) a proposito dell'interpretazione dei risultati di una SCCR di superiorità come SCCR di non inferiorità si afferma: "... in any superiority trial where non-inferiority may be an acceptable outcome for licensing purposes, it is prudent to specify a non-inferiority margin in the protocol in order to avoid the serious difficulties that can arise from later selection". Ciò significa che occorre definire preliminarmente il margine di non inferiorità entro il quale l'effetto del nuovo trattamento può essere considerato uguale a



**Figura 1.** Regola decisionale adottata nella sperimentazione clinica controllata randomizzata GUSTO V per la tipologia superiorità (a) e non inferiorità (b), dove “R” rappresenta il trattamento con reteplase ed “A” il trattamento con abciximab. La regione tratteggiata rappresenta la regione di accettazione dell’ipotesi di interesse clinico ( $\rho_R$  = rischio relativo vero).

quello del controllo attivo, per evitare che il margine stesso venga definito successivamente alla conoscenza dei dati della SCCR con lo scopo di considerare non inferiore il risultato ottenuto.

Inoltre nel paragrafo IV.2.3 del documento citato si legge: “In a superiority trial the full analysis set, based on ITT (“intention-to-treat”) principle is the analysis set of choice ... In a non-inferiority trial the full analysis set and the PP (“per protocol”) analysis have equal importance and their use should lead to similar conclusions for a robust interpretation” (vedi glossario).

Nell’analisi per *intenzione al trattamento* ogni paziente viene considerato appartenente al gruppo del trattamento assegnatogli per randomizzazione (per esempio nuovo trattamento) anche se per motivi, i più disparati, il paziente può non aver completato il trattamento come da protocollo o addirittura può aver ricevuto erroneamente il trattamento alternativo (controllo attivo). Deviazioni dai criteri di trattamento previsti dal protocollo tendono a ridurre la *sensibilità* della sperimentazione portando più facilmente alla conclusione di *non differenza* tra gli effetti dei due trattamenti. Ne consegue che il giudizio di non inferiorità ne rimane avvantaggiato e, in particolare, anche nel caso più rilevante in cui il nuovo trattamento sia realmente meno efficace del controllo attivo. Per ovviare a questo inconveniente, il documento di EMEA suggerisce di eseguire un’ulteriore analisi dei dati *per protocollo* e cioè considerando i soli pazienti che effettivamente hanno ricevuto ambedue i trattamenti, nella quantità e modalità esattamente previste dal protocollo. I risultati di queste due analisi nelle SCCR di non inferiorità devono essere coerenti. Nella SCCR GUSTO V gli autori predefiniscono il margine di non inferiorità, ma non fanno alcun cenno ai risultati ottenuti con l’analisi *per protocollo*. Inoltre gli autori stessi pur non scostandosi dalle succitate direttive di EMEA, sembrano partire con l’idea che il loro trattamento A + R potrebbe avere o una differenza relativa inferiore all’ipotizzato 15%, ma > 0%, o presen-

tare una mortalità addirittura superiore a quella del controllo attivo (R), tant’è che per scopi di “licensing” prevedono contemporaneamente allo studio di superiorità l’eventuale passaggio allo studio di non inferiorità. Ma di fronte a questa eventualità, qual è la reazione del lettore cardiologo tenuto conto delle considerazioni etiche sugli studi di non inferiorità formulate nella seconda nota? D’altra parte rimanendo nell’ambito delle SCCR di superiorità il cardiologo potrebbe infine porsi il quesito: quale dimensione dello studio sarebbe stata necessaria per mettere in luce, a parità di potenza (80%) e di confidenza (95%), una riduzione relativa inferiore al previsto 15%? La risposta è nella seguente tabella:

$\delta_R$	N.
15%	16 600
10%	37 464
7.5%	67 418
5%	153 518

SCCR di superiorità di tali dimensioni sarebbero forse risultate non realizzabili dal punto di vista della fattibilità dello studio, di cui l’aspetto economico è senz’altro una parte rilevante.

### Riassunto

Scopo di questa nota statistica, la quarta della serie, è descrivere e discutere criticamente la sperimentazione clinica controllata randomizzata (SCCR) GUSTO V, che costituisce un tentativo di combinare in una sola sperimentazione le due tipologie di SCCR di superiorità e di non inferiorità. In questa nota si descrive il percorso logico che gli autori hanno seguito per la pro-

grammazione della SCCR sia di superiorità che di non inferiorità e se ne presentano i risultati. Inoltre si mettono in luce i possibili aspetti critici di una SCCR programmata in tal modo e si chiede al lettore di riflettere a riguardo.

*Parole chiave:* Analisi per protocollo; Intenzione al trattamento; Sperimentazioni cliniche controllate randomizzate di non inferiorità; Sperimentazioni cliniche controllate randomizzate di superiorità.

## Glossario

- Analisi sull'insieme completo di dati ("full analysis set"): l'insieme dei soggetti che è più vicino possibile all'ideale implicito nel principio dell'intenzione al trattamento. È derivato dall'insieme di tutti i soggetti randomizzati con un'eliminazione minima e giustificata dei soggetti.
- Principio dell'intenzione al trattamento ("intention-to-treat principle"): il principio che asserisce che l'effetto di una strategia di trattamento può essere meglio valutato in base all'intenzione di trattare un soggetto (cioè il regime di trattamento raccomandato) piuttosto che al trattamento realmente somministrato. Ciò ha come conseguenza che i soggetti assegnati ad un gruppo di trattamento dovrebbero essere seguiti, valutati ed analizzati come componenti di quel gruppo indipendentemente dalla loro aderenza al regime di trattamento previsto.
- Insieme dei soggetti trattati secondo protocollo ("per protocol set"), casi validi, campione di efficacia: l'insieme dei dati generati dal sottoinsieme dei soggetti che hanno aderito al protocollo in modo tale da garantire che questi dati verosimilmente esibiscano gli effetti del trattamento, secondo il modello scientifico sottostante. L'aderenza al trattamento riguarda considerazioni relative all'esposizione al trattamento, alla disponibilità del-

le rilevazioni e all'assenza di importanti violazioni del protocollo.

Da EMEA (The European Agency for the Evaluation of Medicinal Products) - ICH Topic E 9 Statistical Principle for Clinical Trial - Note for guidance on statistical principles for clinical trials. CPMP/ICH/363/96.

## Bibliografia

1. Marubini E, Reina G. Note statistiche. Misure di effetto assolute e relative. Ital Heart J Suppl 2004; 5: 466-71.
2. Marubini E, Rebora P, Reina G. Note statistiche. Dalle sperimentazioni cliniche di superiorità alle sperimentazioni cliniche di non inferiorità: un salto nel vuoto? Ital Heart J Suppl 2004; 5: 712-9.
3. Topol EJ, for the GUSTO V Investigators. Reperfusion therapy for acute myocardial infarction with fibrinolytic therapy or combination reduced fibrinolytic therapy and platelet glycoprotein IIb/IIIa inhibition: the GUSTO V randomized trial. Lancet 2001; 357: 1905-14.
4. The Global Use of Strategies to Open Occluded Coronary Arteries (GUSTO III) Investigators. A comparison of reteplase with alteplase for acute myocardial infarction. N Engl J Med 1997; 337: 1118-23.
5. An international randomized trial comparing four thrombolytic strategies for acute myocardial infarction. The GUSTO Investigators. N Engl J Med 1993; 329: 673-82.
6. Randomised trial of intravenous streptokinase, oral aspirin, both, or neither among 17 187 cases of suspected acute myocardial infarction: ISIS-2. ISIS-2 (Second International Study of Infarct Survival) Collaborative Study. Lancet 1988; 2: 349-60.
7. The European Agency for the Evaluation of Medicinal Products (Committee for Proprietary Medicinal Products). Points to consider on switching between superiority and non-inferiority. CPMP/EWP/482/99.