

Note statistiche. Dalle sperimentazioni cliniche di superiorità alle sperimentazioni cliniche di non inferiorità: un salto nel vuoto?

Ettore Marubini, Paola Rebora, Giuseppe Reina

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

Key words:

Equivalence randomized controlled clinical trials; Margin of clinical tolerance; Non-inferiority randomized controlled clinical trials.

In these “statistical notes”, equivalence and non-inferiority randomized controlled clinical trials (RCCT) are considered. Equivalence trials are designed to confirm the absence of a meaningful difference between the effect of two treatments. Non-inferiority trials are designed to prove that the new treatment is no less effective than an existing one: it may be more effective or it may have a similar effect. In this note the attention is addressed to suitable criteria for the choice of the tolerance margin ϵ , i.e. the largest difference which is clinically acceptable, so that a difference bigger than that would matter in practice. In particular, the procedures for the determination of the margin ϵ , used by the authors of the non-inferiority RCCT COBALT and INJECT, are presented and discussed in detail. The ethical implications of equivalence and non-inferiority RCCT are here considered and the reader is repeatedly invited to consider the appropriateness of the basic arguments asserted by the supporters of this kind of studies.

(Ital Heart J Suppl 2004; 5 (9): 712-719)

© 2004 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 26 maggio 2004; nuova stesura il 29 luglio 2004; accettato il 30 luglio 2004.

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 precedente nota statistica¹ sono state trattate e descritte le sperimentazioni cliniche controllate randomizzate (SCCR) di superiorità le quali, come è noto, mirano ad accertare l'eventuale differenza esistente tra due diversi trattamenti terapeutici della medesima affezione e specificamente a verificare se, e in quale misura, un trattamento possa essere considerato superiore ad un altro. In particolare nella SCCR GISSI² si confrontò la riduzione di mortalità indotta dalla streptochinasi (SK) con quella indotta in un gruppo di riferimento (controllo attivo senza trombolitici) e nella SCCR GUSTO III³ si confrontò la riduzione di mortalità indotta da reteplase (nuovo trattamento) con quella indotta da alteplase (controllo attivo).

Poiché negli ultimi 20 anni si è assistito, in cardiologia, ad un sensibile aumento degli studi di equivalenza e di non inferiorità⁴, nella presente nota si affronteranno le problematiche relative a queste tipologie di studi clinici.

Le sperimentazioni cliniche controllate randomizzate di equivalenza e di non inferiorità

Secondo il documento CPMP/EWP/482/99⁵, le SCCR di equivalenza sono pro-

grammate “... to confirm the absence of a meaningful difference between treatments ... A margin of clinical equivalence ($-\epsilon$, $+\epsilon$) is chosen by defining the largest difference that is clinically acceptable, so that a difference bigger than this would matter in practice”.

Analogamente a quanto sviluppato nella SCCR di superiorità, anche in questo caso il fulcro dell'analisi statistica è rappresentato dall'intervallo di confidenza della vera differenza tra l'effetto dei due trattamenti. Come illustrato in figura 1a, la vera differenza tra la mortalità μ_E del trattamento in esperimento (E) e la mortalità μ_C del suo controllo attivo (C) è riportata sull'asse delle ascisse.

Un giudizio di equivalenza tra i due farmaci impiegati può essere formulato solo se l'intervallo di confidenza della vera differenza è incluso interamente tra le due soglie $-\epsilon$, e $+\epsilon$.

Negli studi di non inferiorità, che sono assai più frequenti di quelli di equivalenza, il ricercatore desidera dimostrare che l'effetto del nuovo trattamento non è peggiore di quello del suo controllo attivo: tale effetto può essere superiore o anche di poco inferiore a quello del controllo attivo, purché la differenza tra i due sia inferiore a una predefinita soglia di tolleranza ϵ . Al fine di consentire un giudizio di non inferiorità del

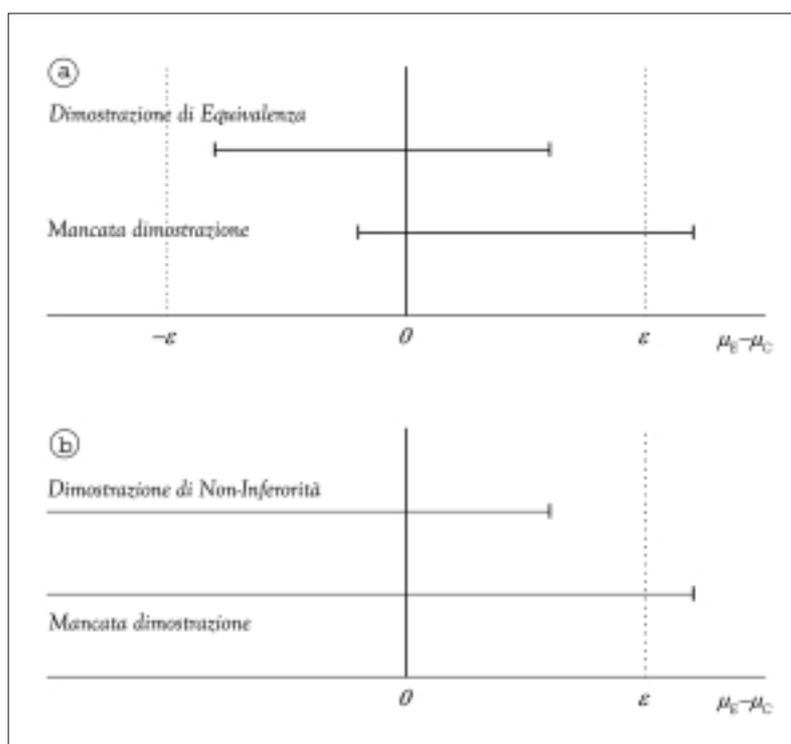


Figura 1. Intervalli di confidenza nell'analisi dei risultati di sperimentazioni cliniche controllate randomizzate di equivalenza (a) e di non inferiorità (b). Sull'asse delle ascisse è riportata la vera differenza $(\mu_E - \mu_C)\%$.

nuovo trattamento rispetto al controllo attivo, in questo caso, come illustrato in figura 1b, è sufficiente che l'estremo superiore dell'intervallo di confidenza della vera differenza $\mu_E - \mu_C$ giaccia interamente alla sinistra della soglia di tolleranza ε .

Di fatto, per quanto consta agli estensori di questa nota, in cardiologia sono oggi utilizzati solo studi di non inferiorità i quali, molto spesso, sono definiti impropriamente studi di equivalenza.

È immediato osservare che la programmazione di uno studio di non inferiorità comporta la rinuncia della ricerca di un trattamento migliore a favore dell'accettazione di un trattamento "simile a", o "non peggiore di" quelli già a disposizione del clinico.

A giustificazione di questa rinuncia, i promotori di SCCR di non inferiorità frequentemente sostengono che i clinici necessitano di farmaci dotati di eguale efficacia ma con minor incidenza di eventi avversi, trascurando di osservare che tale vantaggio è raramente, se non mai, verificabile con le dimensioni degli studi di fase III di corrente uso in cardiologia.

A questo punto il lettore cardiologo avrà realizzato che l'aspetto cruciale di una SCCR di non inferiorità è rappresentato dalla scelta della soglia di tolleranza ε e ciò è tanto più vero quanto più la risposta di interesse clinico è costituita da un evento grave o addirittura fatale. Perciò nel prosieguo l'attenzione sarà rivolta alla valutazione dei criteri di scelta della soglia di tolleranza ε adottati in SCCR pubblicate in riviste di grande prestigio.

La sperimentazione clinica controllata randomizzata di non inferiorità COBALT

La SCCR COBALT⁶ si prefisse di saggiare l'ipotesi che "... *double-bolus alteplase might be at least as effective as the accelerated infusion as used in the GUSTO I study*". La somministrazione di alteplase in infusione continua costituisce quindi il controllo attivo (C) contro cui saggiare l'effetto di alteplase doppio bolo (trattamento sperimentale, E).

Dalla sezione "Endpoint and statistical analysis" si ricava: "*Double-bolus alteplase was to be considered at least equivalent to accelerated infusion if the upper boundary of the one-sided 95% confidence interval of the difference in 30-day mortality did not exceed 0.40 percentage point*".

Ma qual è la ragione che ha indotto gli autori a scegliere $\varepsilon = 0.40\%$?

Obiettivo della SCCR COBALT era quello di confrontare l'effetto di due modalità di somministrazione dello stesso principio attivo. Quindi appare plausibile che gli autori si attendessero un effetto di alteplase a doppio bolo sovrapponibile a quello di alteplase in infusione continua.

L'effetto di alteplase in infusione continua era stato studiato nella SCCR GUSTO I⁷ che aveva posto in evidenza una mortalità $m_C = 6.3\%$; inoltre si era osservato un beneficio dell'1% di alteplase in infusione continua rispetto a SK (mortalità $m_{SK} = 7.3\%$).

La differenza stimata $m_{SK} - m_C = 1\%$ permette di calcolare l'intervallo di confidenza della vera differenza $\mu_{SK} - \mu_C$ che risulta essere pari a $[0.4\%, 1.6\%]$, dove 0.4% rappresenta il beneficio *minimo* di alteplase in infusione continua rispetto a SK; l'attenzione si deve quindi concentrare su quest'ultimo valore, in quanto esso rappresenta una stima *conservativa* dell'effetto di alteplase in infusione continua rispetto a SK.

L'obiettivo che ci si propone ora è quello di fornire uno scenario riguardo alla differenza $\mu_E - \mu_C$, vera ed ignota differenza tra le mortalità di alteplase doppio bolo ed alteplase in infusione continua, al fine di giustificare la scelta di $\varepsilon = 0.4\%$ come soglia di tolleranza clinica. A tale scopo ci si avvarrà dell'evidenza, storicamente acquisita, rappresentata dal *minimo* beneficio di alteplase in infusione continua rispetto a SK ($\mu_{SK} - \mu_C = 0.4\%$) e di una serie di congetture relative alla differenza $\mu_{SK} - \mu_E$, vera differenza tra le mortalità di SK ed alteplase a doppio bolo. Il motivo della serie di congetture su $\mu_{SK} - \mu_E$, risiede nel fatto che nella SCCR COBALT non fu presa in considerazione la possibilità di saggiare, in un terzo gruppo, l'effetto della SK.

In figura 2 sono mostrati tre pannelli relativi alle seguenti differenze di mortalità:

$$^{(1)} \mu_{SK} - \mu_C, \quad ^{(2)} \mu_{SK} - \mu_E, \quad ^{(3)} \mu_E - \mu_C, \quad = (\mu_{SK} - \mu_C) - (\mu_{SK} - \mu_E).$$

Nel primo pannello è riportato l'intervallo di confidenza ottenuto nella SCCR GUSTO I per la vera differenza $\mu_{SK} - \mu_C$.

Nel pannello 2 sono mostrati quattro ipotetici intervalli di confidenza delle vere differenze $\mu_{SK} - \mu_E$.

Nel caso (a) si ipotizza che l'effetto di alteplase a doppio bolo sia esattamente sovrapponibile a quello di alteplase in infusione continua: la differenza tra i limiti inferiori dei due intervalli di confidenza è nulla e si proietta nel punto (a) del terzo pannello; ciò vale anche per la differenza tra i due limiti superiori e quella tra qualsiasi altra coppia omologa di punti dell'intervallo.

Nel caso (b) si ipotizza che il trattamento alteplase a doppio bolo, pur rimanendo più attivo di SK, sia leggermente meno attivo del trattamento di alteplase in infusione continua e la corrispondente differenza $\mu_E - \mu_C$ è proiettata nel punto (b) ($0 < (b) < 0.4\%$).

Nel caso (c) si ipotizza che l'estremo inferiore dell'intervallo di confidenza di $\mu_{SK} - \mu_E$ coincida esattamente con lo zero, il che implicherebbe che il trattamento con alteplase a doppio bolo non possa essere ritenuto più attivo di SK; la proiezione della corrispondente differenza tra i due limiti inferiori coincide con il punto (c) che assume esattamente il valore $\varepsilon = 0.4\%$.

Infine, nel caso (d) si ipotizza che l'estremo inferiore dell'intervallo di confidenza della differenza $\mu_{SK} - \mu_E$ cada sotto lo zero, il che implicherebbe che il trattamento con alteplase a doppio bolo sia meno attivo di quello con SK; nel terzo pannello, questo caso è rappresentato dal punto (d) che risulta $> 0.4\%$.

L'andamento dei punti (a), (b), (c), (d), ottenuto immaginando una continua crescita della mortalità di alteplase a doppio bolo, mostra come $\mu_E - \mu_C = 0.4\%$ sia

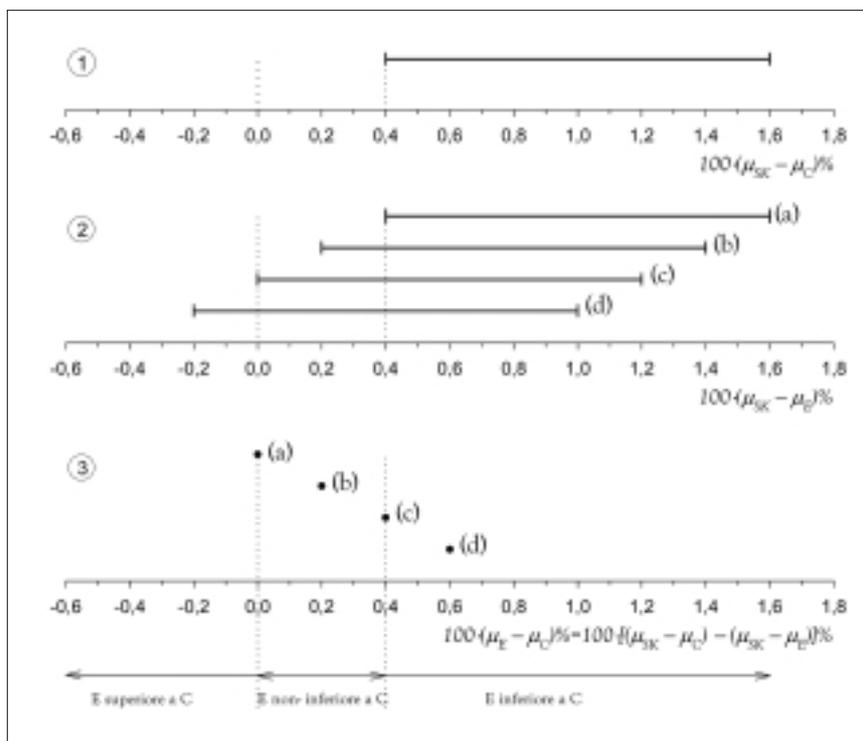


Figura 2. Scenario per l'individuazione del margine di tolleranza adottato nella sperimentazione clinica controllata randomizzata COBALT.

il massimo valore possibile compatibile con l'ipotesi di non inferiorità di alteplase a doppio bolo rispetto ad alteplase in infusione continua e questo costituisce il limite superiore dell'intervallo di confidenza della vera differenza $\mu_E - \mu_C$, come affermato dagli autori di COBALT e sopra riportato.

L'iter logico sopra descritto è così sintetizzato dagli autori della SCCR COBALT: "This value [0.4%] represented the lower 95% confidence limit of the 1% absolute difference in 30-day mortality between an accelerated infusion of alteplase and SK that was observed in the GUSTO I trial".

Dalla sezione "Results" della SCCR COBALT si apprende che $m_E = 7.98\%$ e $m_C = 7.53\%$, con una differenza $m_E - m_C = 0.45\%$.

Si può calcolare che "For the primary endpoint a one-sided 95% upper boundary ..." risulta 1.49%. Poiché quest'ultimo è superiore alla soglia ε prefissata, è ragionevole concludere con gli autori che "... the two treatments were not shown to be equivalent".

La soglia di non inferiorità è presentata nella SCCR COBALT in termini assoluti. Come negli studi di superiorità¹ il clinico potrebbe preferire alla differenza assoluta di effetto tra due trattamenti la differenza relativa, così negli studi di non inferiorità il clinico potrebbe essere facilitato nell'interpretazione della soglia ε dalla formulazione di quest'ultima in termini relativi, precisamente:

$$\frac{\varepsilon}{m_C} = \frac{0.4\%}{6.3\%} = 6.35\%.$$

Quindi, nella SCCR COBALT, gli autori erano disposti a perdere il 6.35% del beneficio relativo indotto da alteplase in infusione continua rispetto a SK pur di somministrare il farmaco sotto forma di doppio bolo.

Il lettore ritiene tollerabile una perdita di questa entità? Noti, peraltro, che si tratta di una perdita relativa tra due modalità di somministrazione dello stesso principio attivo.

La sperimentazione clinica controllata randomizzata di non inferiorità INJECT

La SCCR INJECT⁸ si propose di saggiare l'effetto di reteplase a doppio bolo nei confronti di SK in pazienti con infarto miocardico acuto.

Nella sezione "Statistical methods" si legge: "Estimates of 90 min coronary-artery patency from a previous study together with the results of the GUSTO I angiographic substudy suggested that 35-day mortality for reteplase would be 1% or more lower than that of SK, in which case the center of the confidence interval would be near -1%. If we assume, further, that the treatments could still be equivalent if reteplase was not more than 1% worse than SK, then a sample size must be

chosen to ensure that the confidence interval does not extend as far as +1%".

La lettura dei risultati della SCCR GUSTO I porta qui all'individuazione di una soglia di non inferiorità ($\varepsilon = 1\%$) maggiore di quella utilizzata dalla SCCR COBALT!

Quale base razionale fu di supporto a tale scelta?

Si ricordi che la riduzione di mortalità di alteplase in infusione continua rispetto a SK, dalla SCCR GUSTO I è:

$$m_{\text{alteplase}} - m_{\text{streptochinasi}} = -1\%.$$

Questo valore empirico diviene per gli autori della SCCR INJECT il punto di partenza per ricavare la soglia, assumendo che la differenza tra le mortalità di reteplase e SK sia esattamente uguale a quella osservata dalla SCCR GUSTO I tra alteplase in infusione continua e SK. Gli autori si aspettano quindi che questo -1% diventi il centro del futuro intervallo di confidenza che calcoleranno al termine dello studio. Ma il limite superiore di questo intervallo che valore massimo può assumere?

Gli autori giudicano tollerabile un peggioramento di reteplase rispetto a SK pari all'1% senza darne alcuna giustificazione e quindi fissano arbitrariamente la soglia $\varepsilon = +1\%$. Peraltro essi sentono la necessità di ancorare la scelta di ε alla conoscenza storica dell'effetto di SK, controllo attivo, nei confronti del placebo.

Ciò comporta implicitamente l'ammissione che il disegno di uno studio di non inferiorità, in cui l'effetto del nuovo trattamento (E) è confrontato solo con l'effetto del controllo attivo (C), sia incompleto: manca, infatti il gruppo di trattamento con placebo. D'altro lato sarebbe non etico attivare quest'ultimo, poiché è già noto il beneficio del controllo attivo e d'altro lato ancora si deve considerare che la differenza di effetto tra controllo attivo e placebo può variare da studio a studio e quindi è importante conoscere dove si collochi sulla scala degli effetti veri il nuovo trattamento rispetto sia al controllo attivo sia al placebo, come se fossero saggiati nella stessa SCCR.

Ne emerge l'opportunità di sviluppare una procedura ipotetica, per la scelta del margine ε , articolata su tre livelli come riportato in figura 3⁹ e precisamente:

- *primo*: nelle SCCR di non inferiorità si deve avere una ragionevole fiducia che il controllo attivo darebbe un risultato superiore al placebo se quest'ultimo fosse utilizzato. L'uso dei risultati di precedenti SCCR spesso permette di verificare tale assunto; debbono essere disponibili dati storici con cui si sia stabilito che l'effetto del controllo attivo (C) sia superiore a quello del placebo (P). Inoltre, e questo è un aspetto fortemente critico, ci si deve basare su un assunto molto forte, l'assunto di costanza degli effetti, cioè che la differenza di effetto tra controllo attivo e placebo storicamente osservata, persista nella nuova SCCR di non inferiorità (in cui solo gli effetti del nuovo trattamento e del controllo atti-

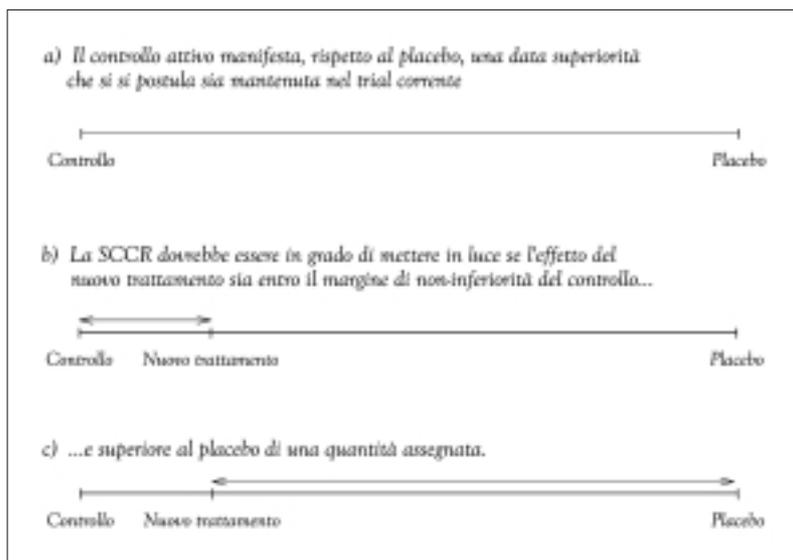


Figura 3. Punti salienti per la determinazione del margine di non inferiorità secondo D'Agostino et al.⁹. SCCR = sperimentazione clinica controllata randomizzata.

vo sono saggiati) qualora il placebo costituisse il trattamento per il terzo gruppo della stessa [a] nella figura 3];

- *secondo*: la SCCR di non inferiorità dovrebbe dimostrare che l'effetto del nuovo trattamento (E) è entro il margine di non inferiorità rispetto al controllo attivo (C) e tale margine dovrebbe essere tollerabile dal punto di vista clinico [b] nella figura 3];
- *terzo*: è quindi necessario usare i dati del confronto fra il nuovo trattamento (E) e il controllo attivo (C) in connessione con i dati storici del confronto del placebo (P) versus il controllo attivo (C). Questa fase comporta il confronto con il *placebo putativo* (storico). Nell'ambito di questo livello è spesso necessario stabilire non solo che il nuovo trattamento (E) sia superiore al placebo (P), ma addirittura che *esso preservi almeno una certa frazione di superiorità del controllo attivo rispetto al placebo* [per esempio, 80% come in c) di figura 3].

Ritornando al paragrafo “Statistical methods”, gli autori della SCCR INJECT informano che: “A meta-analysis of 13 placebo-controlled trials done before this study gave an overall estimate of the benefit of thrombolysis over placebo of 2.7% with the lower limit of the 90% confidence interval being 2.1%”.

Le precedenti affermazioni si traducono in uno schema decisionale in termini di intervalli di confidenza, quale quello riportato in figura 4.

Nel punto $m_{reteplase} - m_{streptochinasi} = 0$, reteplase ha la stessa efficacia di SK; quanto più ci si muove alla destra di zero tanto più l'effetto di reteplase è inferiore a quello di SK fino a raggiungere, nel punto 2.1%, lo stesso effetto del placebo. La zona di non inferiorità è costituita da tutti i valori < 1%, soglia prestabilita arbitrariamente dagli autori.

Il lettore dovrebbe ora essere in grado di rispondere al quesito: quale perdita di beneficio della SK rispetto

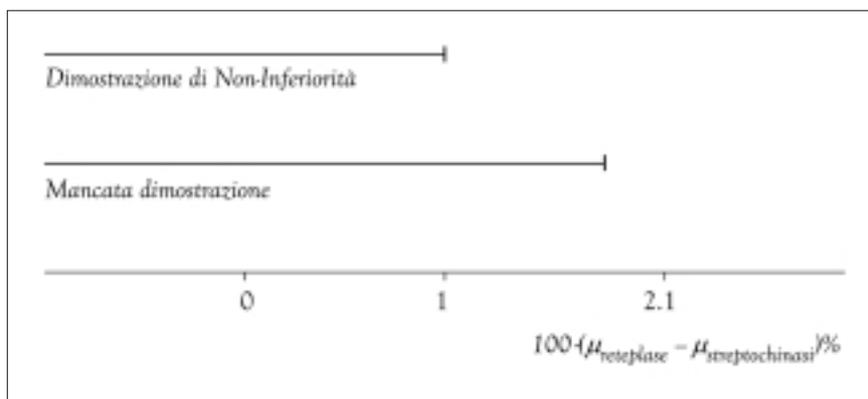


Figura 4. Regola decisionale adottata dalla sperimentazione clinica controllata randomizzata INJECT. Sull'asse delle ascisse è riportata la vera differenza $100 \cdot (\mu_{reteplase} - \mu_{streptochinasi})\%$. Il margine di tolleranza è fissato al valore $\epsilon = 1\%$

al placebo comporta la scelta di $\varepsilon = 1\%$? A tal fine è sufficiente calcolare il rapporto

$$\frac{\varepsilon}{2.1\%} = \frac{1\%}{2.1\%} = 47.6\%.$$

La frazione di beneficio preservata è quindi $100\% - 47.6\% = 52.4\%$.

A prima vista, il margine di non inferiorità $\varepsilon = 1\%$ può apparire una quantità piccola; ma se si considera che a questo corrisponde una perdita di beneficio della SK rispetto al placebo pari al 47.6% può, questa quantità, considerarsi davvero clinicamente tollerabile?

Nella sezione “Results” si legge: “270 (9.02%) randomised patients had died by 35 days in the reteplase group compared with 285 (9.53%) in the SK group, a difference of -0.51% with a 90% confidence interval of -1.74% to 0.73% ... As the confidence interval includes 0, this study has not demonstrated that reteplase is superior to SK, but that it is equivalent”.

La non inferiorità è dimostrata e, *a posteriori*, dal valore riportato dell’estremo superiore dell’intervallo di confidenza della differenza $\mu_{reteplase} - \mu_{streptochinasi}$ (0.73%) si evince che l’eventuale uso di reteplase può comportare una perdita fino al 34.8% del beneficio di SK rispetto al placebo

$$\left(\frac{0.73\%}{2.1\%} = 34.8\% \right).$$

Dalla metanalisi precedentemente citata, risultava che il beneficio del trattamento con SK rispetto al placebo comportava al minimo di salvare 21 pazienti su 1000. L’uso di reteplase invece di SK potrebbe ridurre tale valore fino a: $21 \cdot (1 - 0.348) \cong 14$ e quindi si potrebbero perdere fino a $21 - 14 = 7$ pazienti su 1000 se trattati con reteplase invece che con SK (Tab. I).

Considerazioni finali

L’obiettivo delle SCCR di superiorità, considerate nella prima di queste “note statistiche”, consisteva nel-

l’individuazione di un nuovo trattamento dotato di un effetto superiore a quello del miglior trattamento standard disponibile. Nelle SCCR di non inferiorità si persegue un obiettivo ben più modesto: l’accettazione di trattamenti il cui effetto è simile, od inferiore per non più di ε , a quello del trattamento standard.

Come accennato precedentemente, i fattori di questa tipologia di studi sostengono che il medico pratico necessita di più trattamenti di pari effetto, ma con un’inferiore incidenza di eventi avversi. Sfortunatamente le dimensioni delle SCCR di non inferiorità, che usualmente sono più piccole delle dimensioni delle SCCR di superiorità, non consentono di stimare in maniera affidabile una riduzione di eventi avversi di bassa incidenza.

Appare evidente come tutto ciò abbia delle implicazioni di natura etica, in quanto non si prevede un incremento di beneficio né per i pazienti arruolati nelle SCCR di non inferiorità né per i pazienti futuri.

Su questo argomento un recente articolo di Garattini et al.¹⁰, al quale si rimanda per un’approfondita disamina, ha aperto un ampio dibattito (si veda ad esempio Cesana e Marubini¹¹). Secondo Garattini et al.¹⁰ gli studi di non inferiorità, salvo casi particolari, sono da considerarsi non etici per molteplici ragioni, la prima delle quali attiene alla difficoltà di individuare e giustificare il margine di non inferiorità: “What makes a level of non-inferiority acceptable is whether the research question reflects a real need”.

Appare lecito chiedersi se nel contesto della SCCR INJECT questa asserzione trovi una risposta affermativa.

Ben diversa è la situazione affrontata nella SCCR COBALT dove confrontandosi due modalità diverse di somministrazione dello stesso principio attivo, secondo lo stesso Garattini: “... equivalence or non-inferiority trials are appropriate and do not pose ethical problems”.

Un ulteriore aspetto critico delle SCCR di non inferiorità è rappresentato dalla scelta del controllo attivo. Gli autori della SCCR INJECT adottarono come controllo attivo il capostipite dei trombolitici, la SK, ed evitarono il confronto diretto con alteplase, pur sapen-

Tabella I. Elementi metodologici salienti.

	SCCR COBALT	SCCR INJECT
Definizione	SCCR di equivalenza. Trattasi di fatto di SCCR di non inferiorità	SCCR di equivalenza. Trattasi di fatto di SCCR di non inferiorità
Confronto	Due forme farmaceutiche dello stesso principio attivo (alteplase) <i>Standard:</i> infusione continua <i>Nuova formula:</i> doppio bolo	<i>Standard:</i> streptochinasi <i>Nuovo trattamento:</i> reteplase
Endpoint	Mortalità a 30 giorni	Mortalità a 30 giorni
Margine di tolleranza	$\varepsilon = 0.4\%$	$\varepsilon = 1\%$
Prevista perdita di beneficio	Pari a 6.35%	Pari a 34.8%
Conclusione	La non inferiorità è non dimostrata	La non inferiorità è dimostrata

SCCR = sperimentazione clinica controllata randomizzata.

do dalla SCCR GUSTO I che l'uso di alteplase avrebbe comportato una riduzione di mortalità dell'1% rispetto a SK. Tale confronto, sotto forma indiretta, è peraltro ora possibile considerando congiuntamente i risultati delle SCCR GUSTO I e INJECT in termini di intervalli di confidenza al 95%; questi risultano:

$$\begin{aligned} \mu_{\text{alteplase}} - \mu_{\text{streptochinasi}} &: [-1.59\%; -0.41\%] \\ \mu_{\text{reteplase}} - \mu_{\text{streptochinasi}} &: [-1.98\%; +0.96\%]. \end{aligned}$$

I centri dei due intervalli di confidenza, corrispondenti alla stima puntuale della differenza degli effetti, sono rispettivamente -1% e -0.51%; quest'ultimo valore ci informa che il beneficio di reteplase rispetto a SK è pressoché la metà di quello previsto dagli autori della SCCR INJECT nella programmazione dello studio.

A questo punto il cardiologo che dovesse scegliere tra reteplase e alteplase dovrebbe non avere dubbi, in quanto il primo intervallo, a differenza del secondo, non include lo zero, provando la superiorità di effetto di alteplase rispetto a SK.

Inoltre è opportuno notare che l'ampiezza del secondo intervallo (2.94%) è circa 2.5 volte quella del primo intervallo (1.18%); ciò testimonia come l'effetto di alteplase rispetto a SK sia stimato con maggior precisione di quanto non lo sia l'effetto di reteplase rispetto a SK, coerentemente con l'osservazione già fatta che la numerosità degli studi di superiorità è maggiore di quella degli studi di non inferiorità.

Gli estensori di questa nota ritengono che gli aspetti metodologici e le considerazioni etiche sviluppate da Garattini et al.¹⁰ a proposito dei trial di non inferiorità siano perfettamente condivisibili e sperano che gli esempi adottati possano convincere anche il lettore cardiologo.

Riassunto

Questa "nota statistica" ha come oggetto le sperimentazioni cliniche controllate randomizzate (SCCR) di equivalenza e di non inferiorità. Tali SCCR sono programmate per provare rispettivamente l'assenza di una differenza clinicamente rilevante tra l'effetto di due trattamenti (*equivalenza*) o per dimostrare che un nuovo trattamento *non è peggiore* di un trattamento esistente: l'effetto del nuovo trattamento può essere superiore od anche leggermente inferiore a quello del trattamento esistente purché per non più di una quantità clinicamente tollerabile (*non inferiorità*). Quindi, in entrambi i casi, è necessario introdurre una quantità ϵ , margine di tolleranza, che rappresenti la massima differenza accettabile da parte del clinico, la cui definizione è spesso oggetto di critiche e discussioni. In questa nota, l'attenzione è rivolta ai criteri di scelta di tale margine di tolleranza ed in particolare si descrivono e discutono in dettaglio le procedure usate dagli autori delle SCCR di non inferiorità COBALT ed

INJECT per la determinazione del margine ϵ . Si considerano ripetutamente le implicazioni etiche delle SCCR di equivalenza e non inferiorità e si richiede più volte al lettore di riflettere sulle ragioni addotte dai sostenitori di queste tipologie di studi a favore delle stesse.

Parole chiave: Margine di tolleranza clinica; Sperimentazioni cliniche controllate randomizzate di equivalenza; Sperimentazioni cliniche controllate randomizzate di non inferiorità.

Glossario

- *Margine di tolleranza ϵ .* Definisce la più ampia differenza di effetto, clinicamente tollerabile, per poter considerare due trattamenti equivalenti o non inferiori.
- *Metanalisi.* Analisi statistica di una serie di risultati di precedenti sperimentazioni cliniche controllate randomizzate ovvero di precedenti osservazioni pianificate (studi caso-controllo o studi di coorte) allo scopo di integrare le informazioni per giungere a conclusioni più precise ed affidabili.
- *Preservazione dell'effetto.* Noto l'effetto del controllo attivo rispetto al placebo, per preservazione percentuale dell'effetto si intende la frazione di questo conservata dal nuovo trattamento.
- *Sperimentazioni cliniche controllate randomizzate di equivalenza.* Mirano a confermare l'assenza di differenza tra l'effetto di due trattamenti attivi rispetto a due margini di tolleranza predefiniti dal clinico.
- *Sperimentazioni cliniche controllate randomizzate di non inferiorità.* Mirano a dimostrare che l'effetto di un nuovo trattamento non è peggiore di quello del controllo attivo: tale effetto può essere superiore o anche di poco inferiore a quello del controllo attivo, purché la differenza tra i due sia inferiore a un predefinito margine di tolleranza.

Bibliografia

1. Marubini E, Reina G. Note statistiche. Misure di effetto assolute e relative. Ital Heart J Suppl 2004; 5: 466-71.
2. Gruppo Italiano per lo Studio della Streptochinasi nell'Infarto Miocardico (GISSI). Effectiveness of intravenous thrombolytic treatment in acute myocardial infarction. Lancet 1986; i: 397-402.
3. 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.
4. Barbieri P, Marubini E. Gli studi di equivalenza nelle malattie cardiovascolari. Ital Heart J Suppl 2004; 5: 267-75.
5. 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.
6. The Continuous Infusion versus Double-Bolus Administration of Alteplase (COBALT) Investigators. A comparison of continuous infusion of alteplase with double-bolus administration for acute myocardial infarction. N Engl J Med 1997; 337: 1124-30.
7. The GUSTO Investigators. An international randomized trial comparing four thrombolytic strategies for acute myocardial infarction. N Engl J Med 1993; 329: 673-82.

8. International Joint Efficacy Comparison of Thrombolytics. Randomised double-blind comparison of reteplase double-bolus administration with streptokinase in acute myocardial infarction (INJECT): trial to investigate equivalence. *Lancet* 1995; 346: 329-36.
9. D'Agostino RB, Massaro JM, Sullivan LM. Non-inferiority trials: design concepts and issues - the encounters of academic consultants in statistics. *Stat Med* 2003; 22: 169-86.
10. Garattini S, Bertele V, Li Bassi L. How can research ethics committees protect patients better? *BMJ* 2003; 326: 1199-201.
11. Cesana BM, Marubini E. Further topics for research ethics committees. (letter) <http://bmj.com/cgi/eletters/326/7400/1199#33672>, June 26, 2003.