- Copyright - Il Pensiero Scientifico Editore downloaded by IP 216.73.216.255 Tue, 15 Jul 2025, 23:10:11

Gli studi di equivalenza nelle malattie cardiovascolari

Pietro Barbieri, Ettore Marubini

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

Key words:
Cardiology; Statistics; Trials.

An increasing number of equivalence and non-inferiority trials appeared in recent years among randomized clinical trials, particularly for the evaluation of efficacy of treatments in cardiovascular medicine. These studies are characterized by remarkable methodological issues and important ethical implications. The aim of this review was to underscore some topic methodological problems: setting the equivalence boundary, the special logic used to establish equivalence, and the determinants of sample size. A Medline search was performed in order to identify the main problems in study design and in the interpretation of results of the equivalence and non-inferiority trials in cardiovascular diseases.

(Ital Heart J Suppl 2004; 5 (4): 267-275)

© 2004 CEPI Srl

Ricevuto il 26 gennaio 2004; nuova stesura il 9 marzo 2004; accettato l'11 marzo 2004.

Per la corrispondenza:

Dr. Pietro Barbieri

Via Cufra, 38
20159 Milano
E-mail: pietro.barbieri@fastwebnet.it

Per numerosi decenni i ricercatori clinici hanno disegnato gli studi sperimentali comparativi allo scopo di evidenziare la superiorità di uno o più trattamenti o combinazione di essi rispetto ad uno standard di riferimento. Solamente da pochi lustri, in un numero limitato ma sempre più numeroso di studi sperimentali, è stata saggiata l'affermazione di "equivalenza" tra i trattamenti valutati.

Il capostipite degli studi comparativi, finalizzati ad evidenziare la superiorità di uno o più trattamenti, fu condotto nel 1747 e pubblicato nel 1753 da James Lind[1]: riguardava 12 pazienti affetti da scorbuto suddivisi in sei sottogruppi di trattamento, durante la navigazione della nave Salisbury. È del 1948 la pubblicazione dei risultati della prima sperimentazione clinica controllata randomizzata eseguita per valutare la superiorità della streptomicina rispetto al trattamento tradizionale nella tubercolosi polmonare[2].

La parola chiave "therapeutic equivalency" fu introdotta su Medline nel 1970. La definizione riportata è la seguente: "*The relative equivalency in the efficacy of different modes of treatment of a disease, most often used to compare the efficacy of different pharmaceuticals to treat a given disease*".

Questi riferimenti evidenziano la relativa novità costituita dagli studi di equivalenza rispetto agli studi di superiorità.

L'emergere degli studi di equivalenza appare particolarmente rilevante nell'ambito delle malattie cardiovascolari ed è caratterizzato da alcune criticità, tra cui: 1) definizione della massima differenza di efficacia tollerata tra i trattamenti; 2) modalità di calcolo della dimensione del campione.

Studi di grandi dimensioni sono stati pubblicati recentemente; essi hanno mobilitato ingenti risorse per concludere che i trattamenti valutati erano equivalenti. Ai fini dell'interpretazione dei risultati degli studi sperimentali, è necessario verificare se l'asserzione di equivalenza, o di non inferiorità, di un trattamento rispetto ad un altro corrisponda alla verifica dell'ipotesi di equivalenza o di non inferiorità o non rappresenti invece il fallimento della dimostrazione dell'ipotesi di superiorità. Nel primo caso si ottiene un'effettiva dimostrazione di equivalenza dei trattamenti, per mezzo di uno studio a tal fine programmato; nel secondo caso l'asserzione di equivalenza viene impropriamente dedotta dalla mancata dimostrazione della superiorità di un trattamento rispetto ad un altro, in genere a causa di una dimensione campionaria inadeguata, quindi con una potenza dello studio insufficiente in relazione alla differenza attesa tra i trattamenti.

Gli studi programmati per dimostrare l'equivalenza o la non inferiorità sono caratterizzati da alcune peculiarità che li distinguono dagli studi di superiorità. Uno degli aspetti più rilevanti è costituito dal fatto che nell'ambito della logica induttiva (quella comunemente utilizzata negli studi di superiorità) l'equivalenza esiste solo co-

- Copyright - Il Pensiero Scientifico Editore downloaded by IP 216.73.216.255 Tue, 15 Jul 2025, 23:10:11

me entità teorica; un giudizio sicuro di equivalenza richiederebbe infatti un numero di unità sperimentali infinitamente grande: due farmaci sono sicuramente di eguale efficacia solo se non si riscontra in nessun caso una differenza.

Negli studi di equivalenza e di non inferiorità è necessario determinare la dimensione del campione con un numero definito di unità sperimentali ed è quindi necessario verificare che la differenza tra i trattamenti non sia superiore ad un valore predefinito (Δ), giudicato *a priori* clinicamente rilevante ai fini dell'interscambiabilità dei trattamenti. La logica del test statistico viene capovolta rispetto agli studi di superiorità, in cui la dimostrazione di efficacia di un trattamento rispetto allo standard si ottiene rifiutando l'ipotesi di non differenza tra i trattamenti; al contrario, negli studi di equivalenza o di non inferiorità, la dimostrazione di comparabilità di efficacia si ottiene mediante il rifiuto dell'ipotesi di una differenza superiore a quella predefinita e giudicata clinicamente rilevante *a priori*. Le implicazioni di questo approccio per quanto attiene agli obiettivi e alle specificità di analisi e di interpretazione sono brevemente descritte nei successivi paragrafi; in generale il confronto è effettuato tra uno o più principi attivi contro un principio attivo già conosciuto, tant'è vero che gli studi di equivalenza vengono talora denominati "active control trials".

Questa nota si propone di richiamare alcuni aspetti metodologici critici e specifici degli studi di equivalenza e di non inferiorità, utili ai fini della corretta interpretazione dei loro risultati, soprattutto per quanto attiene al disegno sperimentale, all'analisi dei risultati e alle conclusioni di tali studi nell'ambito delle malattie cardiovascolari. A tal fine è stata preliminarmente effettuata una ricerca su alcune banche dati biomediche orientata ad individuare le voci bibliografiche attinenti alle applicazioni di tale tipologia di studi sperimentali nella valutazione di efficacia dei trattamenti delle malattie cardiovascolari.

Una descrizione degli aspetti più controversi connessi agli studi di equivalenza o di non inferiorità è particolarmente opportuna nel momento attuale, caratterizzato da una continua pressione a contenere i costi sanitari: si assiste alla crescente sollecitazione a utilizzare farmaci generici, ritenuti equivalenti rispetto ai prodotti "di marca" oppure a negare il finanziamento per provvedimenti terapeutici caratterizzati da un rapporto costo-efficacia non favorevole.

Un appropriato giudizio sul rapporto costo-efficacia di trattamenti alternativi si dovrebbe basare su una solida informazione relativa all'equivalenza clinica di trattamenti che hanno costi differenti o su una misura della superiorità dei trattamenti a confronto. In particolare è necessario evitare di procurare danno ai pazienti, come conseguenza di studi mal condotti, proponendo trattamenti clinicamente inferiori ma dichiarati equivalenti o di non riconoscere la superiorità di un trattamento più efficace rispetto allo standard.

Definizioni

Studi di superiorità. Gli studi di superiorità sono progettati per evidenziare una differenza tra trattamenti. Si saggia l'ipotesi (ipotesi nulla: H_0) che i trattamenti abbiano uguale efficacia; se, in base alle evidenze sperimentali, tale ipotesi risulta inverosimile, ad un livello di probabilità convenzionalmente < 5%, si conclude per la superiorità di un trattamento rispetto all'altro (ipotesi alternativa: H_a). Se si decide che l'ipotesi nulla è inverosimile, per il cardiologo è necessario misurare l'entità della "vera" differenza tra i trattamenti, al fine di esprimere un giudizio di rilevanza clinica. La vera differenza tra trattamenti è quella che in teoria si potrebbe ottenere con un campione infinitamente grande, coincidente con l'intera popolazione di riferimento, cioè tutti i pazienti che potrebbero beneficiare del trattamento. In pratica la differenza tra trattamenti viene misurata su un campione, cioè un sottoinsieme limitato e rappresentativo della popolazione di tutti i soggetti possibili.

La vera differenza tra trattamenti può essere espressa calcolando due quantità: una stima puntuale (ad esempio la differenza osservata tra le medie dei singoli gruppi di trattamento) ed una stima intervallare, che definisce un insieme di valori di differenza tra trattamenti, egualmente plausibili alla luce dei risultati sperimentali. Le due seguenti affermazioni, illustrate nella figura 1 e riferite all'esempio A, sono equivalenti:
1) l'intervallo di confidenza al 95% per la differenza *"vera"* tra le medie esclude lo zero; pertanto l'intero insieme dei plausibili valori della differenza tra i trattamenti, che includono la *"vera"* differenza, porta a concludere che un trattamento è superiore rispetto all'altro;
2) c'è una significatività statistica al 5% ($p < 0.05$) nella differenza tra le due medie. Ciò significa che si può affermare che un trattamento è superiore all'altro con una probabilità < 5% di sbagliare. Affermare che $p < 0.05$ equivale ad ammettere una probabilità < 5% che la decisione di dichiarare una differenza tra i trattamenti sia un falso positivo dello studio.

Studi di equivalenza. Gli studi di equivalenza sono progettati allo scopo di verificare l'assenza di una differenza clinicamente rilevante tra i trattamenti. A tal fine è necessario predefinire un valore massimo, clinicamente tollerabile, di differenza tra gli effetti dei trattamenti, cioè un margine di equivalenza clinica. Questa dichiarazione è il passaggio critico di questi studi. Solo dopo aver definito il margine (Δ) di equivalenza clinica è possibile strutturare il test di significatività, procedendo ad un capovolgimento logico rispetto al test statistico tipico degli studi di superiorità: negli studi di equivalenza l'ipotesi nulla (H_0) afferma che i trattamenti differiscono per un valore superiore al margine di equivalenza clinica (Δ). Si propende per l'ipotesi alternativa (H_a), di equivalenza dei trattamenti, quando la probabilità dell'ipotesi nulla è sufficientemente bassa

- Copyright - Il Pensiero Scientifico Editore downloaded by IP 216.73.216.255 Tue, 15 Jul 2025, 23:10:11

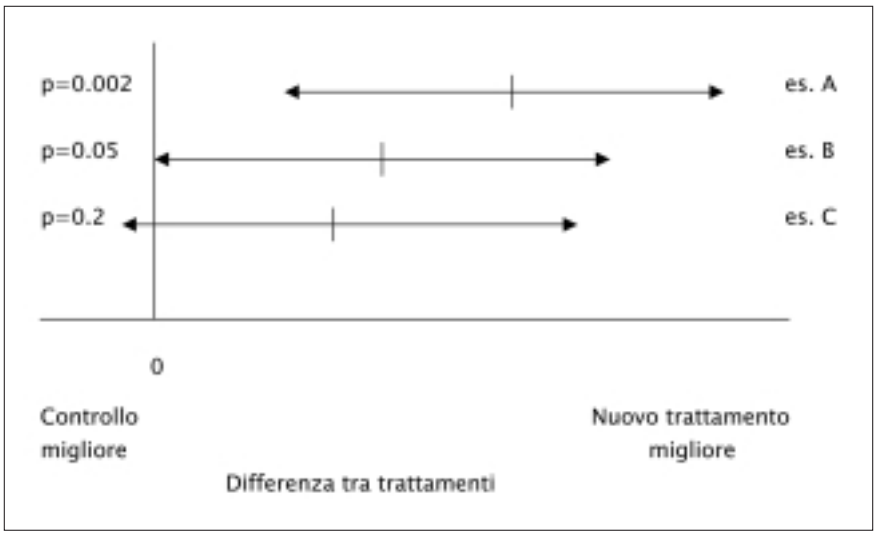

Figura 1. *Relazione tra test di significatività e intervalli di confidenza negli studi di superiorità. Esempio A: dimostrazione di superiorità di un trattamento rispetto ad un altro: l'intervallo di confidenza esclude il valore zero; p < 0.05. Esempio B: caso limite. Esempio C: fallimento della dimostrazione di superiorità: l'intervallo di confidenza include il valore zero; p > 0.05.*

(< 10 o < 5%). La rappresentazione della strategia decisionale mediante l'intervallo di confidenza, in rapporto ai valori di Δ, è illustrata nella figura 2.

Studi di non inferiorità. Nelle sperimentazioni di fase III gli studi di non inferiorità sono più frequenti rispetto a quelli di equivalenza: in questo caso l'obiettivo è dimostrare che un trattamento non è meno efficace dello standard, cioè ha efficacia simile, se non addirittura superiore. Anche in questo caso la rappresentazione della strategia decisionale mediante gli intervalli di confidenza emerge immediatamente: l'interesse è orientato in una sola direzione della differenza di trattamento; l'intervallo di confidenza deve collocarsi interamente a destra del valore di – Δ. Questo aspetto rappresenta il principale elemento di differenziazione degli studi di non inferiorità rispetto a quelli di equivalenza (Fig. 3).

È evidente quanto sia rilevante, ai fini del giudizio finale sull'efficacia di un nuovo trattamento, la scelta del valore di Δ, la differenza di efficacia clinicamente tollerabile. Osserviamo che è sempre possibile definire un appropriato valore di Δ, tale da consentire un giudizio di equivalenza o di non inferiorità, dopo avere preso visione dei risultati, ma ciò invalida tutta la procedura di inferenza statistica. È indispensabile pertanto che il valore di Δ sia definito *a priori*, in base a solide informazioni disponibili al momento della programmazione dello studio, coerenti con l'insieme delle evidenze sperimentali già acquisite: è ragionevole comunque ipotizzare un valore di Δ che sia inferiore alla più piccola differenza tra trattamento standard ("active con-

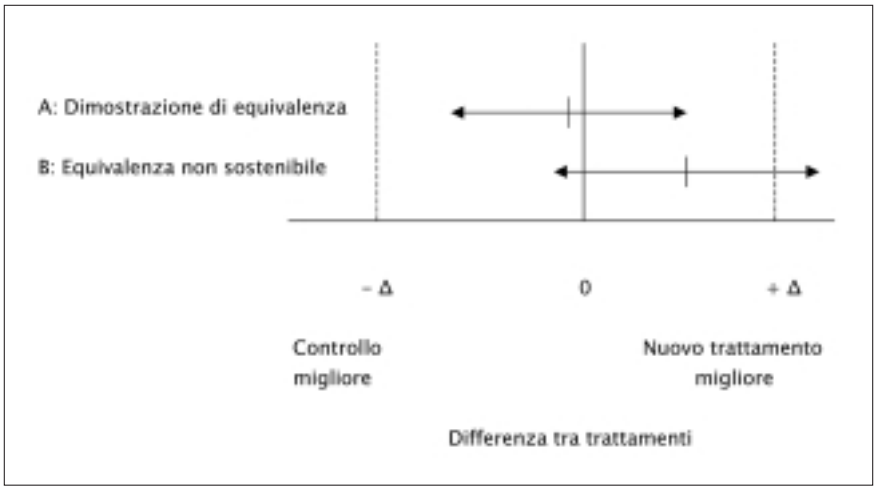

Figura 2. *Studi di equivalenza: strategia decisionale mediante l'uso degli intervalli di confidenza. A: l'intervallo non include Δ; B: l'intervallo include Δ.*

- Copyright - Il Pensiero Scientifico Editore downloaded by IP 216.73.216.255 Tue, 15 Jul 2025, 23:10:11

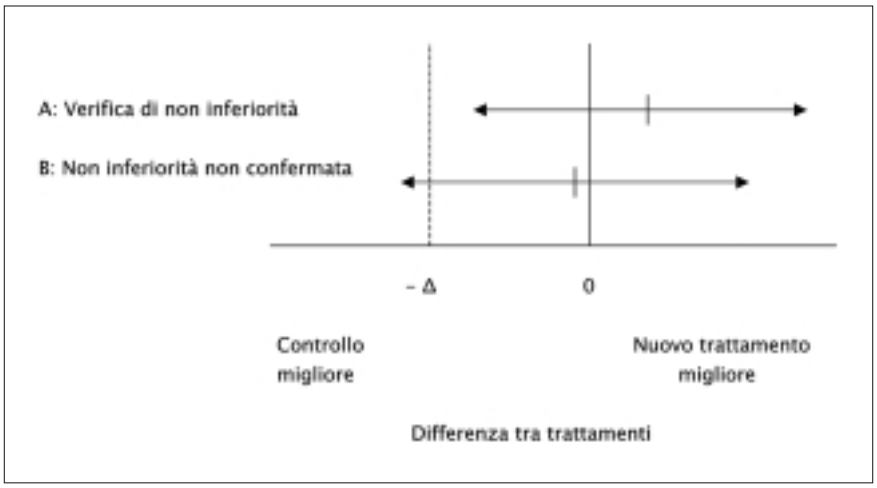

Figura 3. *Studi di non inferiorità: strategia decisionale mediante l'uso degli intervalli di confidenza. A: l'intervallo non include – Δ; B: l'intervallo include – Δ.*

trol") e placebo riscontrata in precedenti studi sperimentali; in ogni caso la determinazione del valore di Δ è un giudizio di rilevanza clinica, della differenza di efficacia massima tollerabile, e non il risultato di un calcolo statistico.

Analogamente devono essere definiti *a priori*: il valore di probabilità utilizzato per la costruzione degli intervalli di confidenza (usualmente 95%) e la potenza attesa dello studio (usualmente non inferiore a 80%), cioè la capacità di giungere ad una corretta conclusione rispetto al giudizio di equivalenza o di non inferiorità.

Un'ulteriore osservazione si riferisce alla modalità di analisi dei dati: nelle sperimentazioni cliniche controllate randomizzate di superiorità è comunemente adottato l'approccio di analisi per "intention to treat" in base al quale, nella valutazione dell'effetto dei trattamenti, tutti i soggetti sono considerati appartenenti al proprio gruppo di randomizzazione, indipendentemente dal fatto che sia stato possibile somministrare a ciascuno il trattamento previsto. Tale approccio si è dimostrato il più robusto rispetto a tutte le possibili distorsioni nell'analisi dei dati nei confronti dell'approccio alternativo ("per protocol"), in base al quale i singoli soggetti vengono valutati in funzione dell'effettivo trattamento seguito. Si preferisce evitare tali distorsioni nell'analisi, al costo di una sottostima della differenza di efficacia teorica tra i trattamenti. Negli studi di equivalenza la situazione è diversa, in quanto la sottostima della differenza potrebbe portare ad un errato giudizio di equivalenza. È consigliabile pertanto analizzare i dati secondo entrambe le modalità (per "intention to treat" e "per protocol") e verificare se si ottengono conclusioni coincidenti[3]; in caso contrario è necessaria un'analisi approfondita dei fattori che hanno prodotto la discrepanza delle evidenze.

Strategia di ricerca su Medline. Al fine di valutare i profili di progettazione e le evidenze sperimentali ottenute dagli studi di equivalenza o di non inferiorità in ambito cardiovascolare, è stata condotta una ricerca su Medline, utilizzando il motore di ricerca PubMed della National Library of Medicine: la ricerca è stata effettuata principalmente utilizzando i MeSH terms (Medical Subject Headings) e parzialmente tramite "free text": il principale termine MeSH utilizzato è "therapeutic equivalency", introdotto nel 1970.

Effettuando una *intersezione logica* tra "therapeutic equivalency" (MeSH) e "clinical trials" (MeSH) si selezionano 290 voci bibliografiche. Effettuando una ricerca mediante "therapeutic equivalency" (MeSH), delimitata dal tipo di pubblicazione "randomized controlled trial" si ottengono 629 voci bibliografiche.

Il differente risultato ottenuto utilizzando concetti analoghi ma qualificatori differenti (esclusivamente MeSH terms vs MeSH terms + qualificatori della tipologia di pubblicazione) dipende dalle specifiche caratteristiche classificatorie degli operatori di Medline. Appare più appropriata la seconda opzione in quanto più pertinente in relazione all'obiettivo della ricerca, focalizzata su "clinical trial" rispetto ad altre tipologie di strumenti informativi (casi clinici, linee guida, editoriali). Un altro motivo per cui la seconda opzione appare più appropriata è costituito dal fatto che conduce ad un insieme più ampio su cui operare una successiva selezione con strumenti più specifici.

È evidente infatti la necessità di non restringere *a priori* l'individuazione delle voci bibliografiche: a tale scopo è stata inoltre effettuata una ricerca non limitata ai termini MeSH, ma estesa anche ai "free text", analogamente a quanto è stato fatto da Greene et al.[4] limitatamente agli anni 1992-1996 per i "clinical trial" in generale.

Effettuando una *unione logica* tra "therapeutic equivalency" (MeSH), equivalence (all fields), noninferiority o non-inferiority (all fields) si ottengono 7502 voci bibliografiche (aggiornamento del 15 settembre 2003).

- Copyright - Il Pensiero Scientifico Editore downloaded by IP 216.73.216.255 Tue, 15 Jul 2025, 23:10:11

Limitando per "randomized controlled trial" si ottengono 1140 voci.

Introducendo tramite *intersezione logica* il termine "cardiovascular diseases" (MeSH) si ottengono 107 voci bibliografiche. Di queste 107 voci sono stati valutati il titolo e l'abstract: in seguito alla loro lettura sono stati successivamente esclusi i riferimenti collegati a studi sperimentali di farmacocinetica, la maggior parte dei quali nel campo dell'ipertensione arteriosa.

L'attenzione si è così concentrata su 28 studi con endpoint principale riconducibile ad una o più variabili dicotomiche: mortalità, eventi cardiaci avversi maggiori, recidiva di eventi tromboembolici[5-32].

Una ricerca integrativa effettuata su Embase non ha modificato il risultato finale, in quanto sono state individuate voci aggiuntive soprattutto orientate agli studi di farmacocinetica.

Metodologia di analisi delle voci bibliografiche. L'intero testo delle voci bibliografiche selezionate è stato valutato in modo strutturato, facendo riferimento a 5 punti qualificanti della sperimentazione:

1) dichiarazione della finalità dello studio. L'esplicita definizione degli obiettivi dello studio, anche in relazione all'attesa superiorità o equivalenza dei trattamenti a confronto, consente di mettere a fuoco in modo univoco tutti gli elementi di progettazione della sperimentazione (numerosità, differenza tra trattamenti ritenuta clinicamente rilevante, modalità di analisi) e favorisce l'interpretazione non equivoca dei risultati;

2) variabile/i di risposta selezionata/e (outcome): consente il giudizio di confrontabilità con altre evidenze sperimentali e il giudizio di importanza clinica;

3) valori di riferimento nel gruppo di controllo (ad esempio la percentuale di eventi nel caso si studi la mortalità) e, ove esistano, stime già disponibili di differenza di efficacia tra i trattamenti valutati. Tali dati consentono di esprimere un giudizio di coerenza del disegno dell'esperimento rispetto alle informazioni già acquisite;

4) determinazione della differenza ritenuta clinicamente rilevante (Δ) ed esplicitazione dei criteri che hanno condotto a tale determinazione. Rappresenta l'elemento critico rispetto alla validità delle conclusioni;

5) aspetti di metodologia statistica: programmazione dello studio e analisi dei risultati. Questa parte della valutazione si riferisce alla definizione di alfa (probabilità dell'errore di primo tipo) e della potenza del test statistico (1 – probabilità dell'errore di secondo tipo). La determinazione della potenza, esplicitata nella parte metodologica o implicita, conseguente alle altre scelte effettuate, ha un peso ancora maggiore negli studi di equivalenza e di non inferiorità, rispetto agli studi di superiorità. L'appropriatezza della dimensione dello studio è una conseguenza dei criteri precedentemente definiti. Il tipo di test statistico utilizzato e la verifica in merito all'affermazione di equivalenza (o alla sua confutazione), se ottenuta mediante un test diretto di equivalenza, relativamente a un Δ predefinito, o alla mancata verifica di superiorità, costituiscono l'elemento conclusivo di questo criterio di valutazione.

Analisi delle voci bibliografiche selezionate. In seguito alla lettura dei titoli e degli abstract sono stati selezionati, per l'analisi dell'intero testo, 28 voci bibliografiche: 6 si riferivano a procedure di rivascolarizzazione coronarica[6-8,10,28,29], 6 alla trombolisi nell'infarto miocardico acuto[11-14,20,24] e 9 all'uso degli antitrombotici[9,16-19,23,25,26,32], utilizzati prevalentemente nella profilassi della trombosi venosa profonda e delle tromboembolie; 7 infine si riferivano a singoli argomenti, ad esempio telemedicina nello scompenso cardiaco, antibiotici nell'endocardite, controllo del ritmo o della frequenza cardiaca nella fibrillazione striale[1,15,21,22,27,30,31].

Rispetto ai 5 criteri di valutazione degli studi precedentemente illustrati, è possibile sintetizzare le seguenti osservazioni:

• l'obiettivo della dimostrazione di equivalenza era esplicitamente dichiarato in 20 studi (71%)[7-14,17-20,23,25-30,32]. In due ulteriori studi l'obiettivo dello studio era dichiarato in modo non univoco, contemplando sia la verifica di superiorità che quella di non inferiorità. In 14 lavori (tra quelli in cui era ben definito l'obiettivo dello studio) compariva un esplicito ed esclusivo riferimento alla valutazione di non inferiorità rispetto ad uno standard di riferimento, a conferma del fatto che negli studi di fase III è questo l'obiettivo prevalente all'interno della famiglia degli studi di equivalenza;

• la definizione della/e variabile/i di risposta è precisa in 26 lavori (93%); in altri due[1,22] sono valutate numerose variabili e non è definita in modo preciso una variabile di risposta principale. Negli studi di rivascolarizzazione prevalgono gli eventi cardiaci avversi maggiori; negli studi sulla trombolisi la mortalità a 30 giorni; negli studi sugli antitrombotici gli eventi tromboembolici combinati, a distanza variabile di tempo;

• la percentuale di eventi nel gruppo di riferimento era dichiarata in 23 casi (82%)[6-9,11-17,19-22,24-29,31,32], ma una stima della differenza di efficacia dei trattamenti era disponibile solo in 11 casi (39%);

• la determinazione della differenza ritenuta clinicamente rilevante per negare l'equivalenza era riportata in 24 casi (86%)[7-21,24-32], ma la lettura trasversale di tali informazioni consente di individuare alcune incongruenze: due studi[13,20] effettuati in anni vicini (INJECT 1993-1994 e COBALT 1995-1996) predefiniscono due valori di Δ differenti, rispettivamente 1 e 0.4%, entrambi avvalendosi di informazioni fornite dallo stesso studio precedente (GUSTO I); il primo fa riferimento al valore centrale dell'intervallo di confidenza e il secondo al limite inferiore dell'intervallo. Queste scelte condizionano nel primo caso un'affermazione di equivalenza dei trattamenti e nel secondo il rifiuto della dimostrazione di non inferiorità, anche se le differenze assolute sono molto simili (rispettivamente 0.51 e 0.44%).

- Copyright - Il Pensiero Scientifico Editore downloaded by IP 216.73.216.255 Tue, 15 Jul 2025, 23:10:11

Rispetto all'ultimo criterio di valutazione dei lavori, si osserva che una dichiarazione esplicita di tutti gli elementi necessari per il calcolo della dimensione campionaria è stata fatta in 15 dei 20 studi che avevano dichiarato esplicitamente come obiettivo la dimostrazione di equivalenza[7-9,11-14,17,19,25,27-30,32]. In questi 15 sono stati eseguiti i pertinenti test per la verifica dell'equivalenza; negli altri 5 la verifica di equivalenza utilizza strumenti non sempre pertinenti.

In 7/28 studi (25%) la dimostrazione di equivalenza è dedotta dalla mancata verifica di superiorità di un trattamento rispetto allo standard: 6 di essi hanno un inizio anteriore al 1996.

I dettagli relativi al calcolo della dimensione del campione e ai valori delle probabilità degli errori di primo e secondo tipo sono riportati nelle tabelle I-III.

Tra i 28 studi analizzati non sempre è dichiarata la modalità di analisi dei dati rispetto a questo criterio, ma in ogni caso prevale il principio per "intention to treat".

Di seguito si riportano alcuni esempi tratti dai 28 lavori selezionati, allo scopo di evidenziare alcuni punti, riportando alcune affermazioni, meritevoli di ulteriori commenti.

Esempio 1. Tratto dallo studio DIRECT[7].

"The aim of this study was to demonstrate therapeutic equivalence". "The null hypothesis was that the 6-month major adverse cardiac event-free survival rate was at least 10% greater with the conventional technique than with direct stenting". "With an estimated 80% 6-month major adverse cardiac event-free survival rate for ... predilation and considering an alfa value of 0.05 and a statistical power of 80%, 394 patients were required". "Outcomes were analyzed according to the intention to treat principle". "Major adverse cardiac event-free survival rate was 87.5% for ... direct stent technique and 85.5% for ... predilation (p = 0.0002 for equivalence)". "Direct stenting is at least equivalent to the standard technique ... This strategy is associated with decreased use of balloons ...".

Commenti. L'obiettivo è formulato in modo chiaro. Gli elementi necessari per il calcolo della dimensione del campione sono esplicitati. Il risultato dello studio è quello atteso; la conclusione è coerente con i risultati e corrisponde ad un uso appropriato degli studi di equivalenza, in quanto consente di dimostrare che una strategia di trattamento meno costosa e meno invasiva ha pari efficacia rispetto alla strategia convenzionale. I punti critici sono rappresentati da: 1) la soglia della differenza clinicamente tollerabile ("major adverse cardiac event-free survival rate" pari a 10%), che appare piuttosto ampia e che, come conseguenza, permette agli autori di eseguire la sperimentazione clinica controllata randomizzata con un numero di pazienti (n = 394) relativamente modesto; 2) modalità di analisi dei dati per "intention to treat".

Esempio 2. Tratto dallo studio GUSTO V[24]. Il punto critico è tutto contenuto nella dichiarazione della finalità dello studio:

"Our aim was to find out whether the combination of half dose reteplase and abciximab would be superior, or not inferior, to reteplase alone ...".

È preferibile individuare un unico obiettivo (superiorità o non inferiorità) altrimenti è difficile evitare di ipotizzare che uno studio programmato per evidenziare la superiorità di un trattamento subisca una trasformazione verso un'ipotesi di non inferiorità in seguito all'analisi dei dati. In particolare la trasformazione da un'ipotesi di superiorità ad un'ipotesi di non inferiorità pone dei problemi in quanto è sempre possibile definire *a posteriori* un valore di differenza clinicamente rilevante (Δ) tale da ottenere un test di equivalenza o non inferiorità statisticamente significativo. Più legittimo appare il percorso inverso, illustrato dal successivo esempio.

Esempio 3. Tratto dallo studio TARGET[8].

"The current trial was designed to test whether the small molecule tirofiban ... was not inferior to the monoclonal antibody abciximab ... in patients who were expected to undergo coronary stenting" (la dimostrazione di non inferiorità si riferiva agli eventi cardiaci avversi maggiori a 30 giorni).

"The upper bound of the 95% confidence interval of the hazard ratio ... had to be less than 1.47, consistent with the preservation of at least 50% in the effect of abciximab as compared with that of placebo in the EPISTENT trial". "The upper bound of the one-sided 95% confidence interval was determined to be 1.51". "Since the non-inferiority of tirofiban as compared with abciximab was not established, we could then assess whether it was superior to abciximab without sta-

Tabella I. Determinazione della probabilità dell'errore di primo tipo.

Probabilità (α) dell'errore di primo tipo	Non determinata	0.1	0.05	0.01
Numero di studi	3	2	22	1

Tabella II. Determinazione della potenza dello studio.

1 - probabilità (β) dell'errore di secondo tipo	Non determinata	0.8	> 0.8
Numero di studi	7	13	8

Tabella III. Determinazione della dimensione del campione.

Dimensione del campione	Non determinata	$< 10^3$	10^3-10^4	$> 10^4$
Numero di studi	3	10	10	5

- Copyright - Il Pensiero Scientifico Editore downloaded by IP 216.73.216.255 Tue, 15 Jul 2025, 23:10:11

tistical penalty". "The assessment of whether abciximab was superior to tirofiban showed a significant difference (hazard ratio 1.26; 95% confidence interval 1.01 to 1.57)". Si noti che si considera la superiorità del trattamento di riferimento rispetto al nuovo trattamento.

"Although the trial was intended to assess the noninferiority of tirofiban as compared with abciximab, the findings demonstrated that tirofiban offered less protection ...".

Commenti. Lo studio ha un unico obiettivo definito. Gli elementi per il calcolo della dimensione del campione sono esplicitati e coerenti con le precedenti evidenze sperimentali. La sperimentazione contraddice l'ipotesi di equivalenza e legittimamente[33] si esegue un test di superiorità che dimostra una significativa differenza tra i trattamenti: in questo caso il margine di equivalenza era stato ovviamente definito *a priori*. Le conclusioni sono coerenti con l'analisi dei dati.

Un ulteriore criterio di valutazione degli studi sperimentali in generale, e degli studi di equivalenza in particolare, è definito da alcune considerazioni di tipo etico[34]. Tra queste le importanti sono le seguenti:

• possibilità di esprimere il margine di equivalenza dal punto di vista etico, soprattutto in relazione ad alcune tipologie di outcome: in base a quali criteri un clinico può determinare la massima differenza di mortalità eticamente tollerabile?

• in relazione al consenso informato per la partecipazione alla sperimentazione: è necessario prospettare ai pazienti che non è atteso alcun vantaggio rispetto al trattamento convenzionale in caso di attribuzione casuale al nuovo trattamento in studio e che comunque sono sempre possibili effetti indesiderati;

• la dimensione del campione in genere è tale che difficilmente sarà possibile evidenziare con sufficiente sensibilità una superiorità rispetto agli effetti collaterali in uno studio dimensionato per verificare l'equivalenza per l'endpoint principale.

Conclusioni

Gli studi di equivalenza hanno un ruolo specifico all'interno degli studi sperimentali: essi sono caratterizzati da alcune peculiarità metodologiche che consistono nel capovolgimento della logica del test statistico, rispetto agli studi di superiorità, e nella necessità di definire *a priori* una differenza clinicamente rilevante, non solo per il calcolo della dimensione del campione, ma anche per la corretta effettuazione della verifica dell'ipotesi.

Ai fini dell'interpretazione dei risultati degli studi sperimentali, è necessario verificare se la dimostrazione di equivalenza, o di non inferiorità, di un trattamento rappresenti il fallimento della dimostrazione di superiorità o la verifica dell'ipotesi di equivalenza o di non inferiorità.

Nel primo caso è implicita l'ammissione di avere sprecato risorse nella programmazione e nella conduzione dello studio sperimentale ma si evita la distorsione connessa alla mancata pubblicazione, da parte dei potenziali autori o degli editori, di rapporti relativi a sperimentazioni con risultati negativi.

Questa evenienza va distinta dalle situazioni in cui lo studio è programmato *a priori* per evidenziare un'equivalenza o una non inferiorità dei trattamenti sperimentali: in tal caso lo studio presuppone obiettivi, disegno sperimentale e modalità di analisi dei dati specifici e distinti da quelli degli studi di superiorità. La dimostrazione di equivalenza rappresenta il risultato atteso e quindi un successo dello studio. Può darsi un terzo caso, nel quale uno studio inizialmente disegnato per dimostrare la superiorità di un trattamento viene successivamente presentato, alla luce dei risultati ottenuti, come uno studio disegnato per dimostrare l'equivalenza, al fine di giustificarne il fallimento come studio di superiorità.

L'ampiezza della differenza ritenuta clinicamente tollerabile è il singolo fattore più importante per la corretta progettazione e interpretazione degli studi di equivalenza o di non inferiorità in quanto è possibile dimostrare che differenti scelte di Δ possano portare a conclusioni contrastanti.

L'obiettivo dello studio deve essere chiaramente formulato: non inferiorità e superiorità sono affermazioni mutuamente esclusive. In seguito all'analisi dei dati è possibile saggiare l'ipotesi di superiorità nel caso in cui sia stata accettata l'ipotesi di non inferiorità, a condizione che sia possibile affermare un identico profilo di sicurezza per i due trattamenti. L'operazione inversa (saggiare l'ipotesi di non inferiorità quando è fallita la dimostrazione di superiorità) appare più problematica in relazione alla necessità di definire *a priori* il margine di equivalenza.

È necessario ricordare che gli studi di equivalenza possono essere considerati appropriati dal punto di vista etico solo in alcuni contesti di riferimento: modalità di somministrazione più facile o trattamento più accettabile (meno invasivo); motivazioni di salute pubblica: più favorevole rapporto costo-efficacia, minore impatto ambientale.

Riassunto

Gli studi di equivalenza e di non inferiorità sono sempre più frequentemente utilizzati nelle sperimentazioni cliniche controllate, in particolare nello studio di efficacia dei trattamenti delle malattie cardiovascolari. Questi studi sono caratterizzati da alcune peculiarità metodologiche e da importanti implicazioni di tipo etico. Questa nota si propone di illustrare i principali punti critici dal punto di vista metodologico, come la definizione della differenza tra trattamenti giudicata clinicamente significativa, le peculiarità che caratterizzano

- Copyright - Il Pensiero Scientifico Editore downloaded by IP 216.73.216.255 Tue, 15 Jul 2025, 23:10:11

la verifica delle ipotesi e le modalità di analisi dei dati e gli elementi necessari al calcolo della corretta dimensione campionaria; presenta i risultati di una revisione bibliografica degli studi di equivalenza e di non inferiorità nelle malattie cardiovascolari, evidenziando le principali criticità riscontrate in merito al disegno dello studio e all'interpretazione dei risultati.

Parole chiave: Cardiologia; Statistica; Trial.

Bibliografia

1. Lind J. A Treatise of the Scurvy. Edinburgh: Sands Murray & Cochran, 1753.
2. Medical Research Council. Streptomycin treatment of pulmonary tuberculosis. BMJ 1948; 2: 769-82.
3. Snapinn SM. Noninferiority trials. Curr Control Trials Cardiovasc Med 2000; 1: 19-21.
4. Greene WL, Concato J, Feinstein AR. Claims of equivalence in medical research: are they supported by the evidence? Ann Intern Med 2000; 132: 715-22.
5. Jenkins RL, McSweeney M. Assessing elderly patients with congestive heart failure via in-home interactive telecommunication. J Gerontol Nurs 2001; 27: 21-7.
6. van den Brand MJ, Rensing BJ, Morel MA, et al. The effect of completeness of revascularization on event-free survival at one year in the ARTS trial. J Am Coll Cardiol 2002; 39: 559-64.
7. Brito FS, Caixeta AM, Perin MA, et al, for the DIRECT Study Investigators. Comparison of direct stenting versus stenting with pre-dilation for the treatment of selected coronary narrowings. Am J Cardiol 2002; 89: 115-20.
8. Topol EJ, Molinterno DJ, Herrmann HC, et al, for the TARGET Investigators. Comparison of two platelet glycoprotein IIb/IIIa inhibitors, tirofiban and abciximab, for the prevention of ischemic events with percutaneous coronary revascularization. N Engl J Med 2001; 344: 1888-94.
9. Merli G, Spiro TE, Olsson CG, et al, for the Enoxaparin Clinical Trial Group. Subcutaneous enoxaparin once and twice daily compared with intravenous unfractionated heparin for treatment of venous thromboembolic disease. Ann Intern Med 2001; 134: 191-201.
10. Baim DS, Cutlip DE, Midei M, et al, for the ASCENT Investigators. Final results of a randomized trial comparing Multi-Link stent with the Palmaz-Schatz stent for narrowings in native coronary arteries. Am J Cardiol 2001; 87: 157-62.
11. Assessment of the Safety and Efficacy of a New Thrombolytic Investigators. Single-bolus tenecteplase compared with front-loaded alteplase in acute myocardial infarction: the ASSENT-2 double-blind randomised trial. Lancet 1999; 354: 716-22.
12. Tebbe U, Michels R, Adgey J, et al. Randomized, double-blind study comparing saruplase with streptokinase therapy in acute myocardial infarction: the COMPASS Equivalence Trial. Comparison Trial of Saruplase and Streptokinase (COMPASS) Investigators. J Am Coll Cardiol 1998; 31: 487-93.
13. 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.
14. 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.
15. Ribera E, Gomez-Jimenez J, Cortes E, et al. Effectiveness of cloxacillin with and without gentamicin in short-term therapy for right-sided Staphylococcus aureus endocarditis. A randomized, controlled trial. Ann Intern Med 1996; 125: 969-74.
16. Julian DG, Chamberlain DA, Pocock SJ. A comparison of aspirin and anticoagulation following thrombolysis for myocardial infarction (the AFTER study): a multicentre unblinded randomised clinical trial. BMJ 1996; 313: 1429-31.
17. Bergmann JF, Neuhart E. A multicentre randomized double-blind study of enoxaparin compared with unfractionated heparin in the prevention of venous thromboembolic disease in elderly in-patients bedridden for an acute medical illness. Thromb Haemost 1996; 76: 529-34.
18. Vives MJ, Hozack WJ, Sharkey PF. Fixed minidose versus-adjusted low-dose warfarin after total joint arthroplasty. A randomized prospective study. J Arthroplasty 2001; 16: 1030-7.
19. Herenberg J, Roebruck P, Heene DL. Subcutaneous low-molecular-weight heparin versus standard heparin and the prevention of thromboembolism in medical inpatients. Haemostasis 1996; 26: 127-39.
20. The INJECT Investigators. Randomised, double-blind comparison of reteplase double-bolus administration with streptokinase in acute myocardial infarction (INJECT): trial to investigate equivalence. International Joint Efficacy Comparison of Thrombolytics. Lancet 1995; 346: 329-36.
21. Gruppo Italiano per lo Studio della Sopravvivenza nell'Infarto Miocardico. GISSI-3: effects of lisinopril and transdermal glyceryl trinitrate singly and together on 6-week mortality and ventricular function after acute myocardial infarction. Lancet 1994; 343: 1115-22.
22. Muller JE, Turi ZG, Pearle DL, et al. Nifedipine and conventional therapy for unstable angina pectoris: a randomized, double-blind comparison. Circulation 1984; 69: 728-39.
23. Planes A. Comparison of antithrombotic efficacy and haemorrhagic side-effects of clivarin versus enoxaparin in patients undergoing total hip replacement surgery. Blood Coagul Fibrinolysis 1993; 4: 33-5.
24. 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 randomised trial. Lancet 2001; 357: 1905-14.
25. The Columbus Investigators. Low-molecular-weight heparin in the treatment of patients with venous thromboembolism. N Engl J Med 1997; 337: 657-62.
26. Kleber FX, Witt C, Vogel G, et al. Randomized comparison of enoxaparin with unfractionated heparin for the prevention of venous thromboembolism in medical patients with heart failure or severe respiratory disease. Am Heart J 2003; 145: 614-21.
27. Black HR, Elliott WJ, Grandits G, et al, for the CONVINCE Study Group. Principal results of the controlled onset verapamil investigation of cardiovascular end points (CONVINCE) trial. JAMA 2003; 289: 2073-82.
28. Lincoff AM, Bittl JA, Harrington RA, et al, for the REPLACE-2 Investigators. Bivalirudin and provisional glycoprotein IIb/IIIa blockade compared with heparin and planned glycoprotein IIb/IIIa blockade during percutaneous coronary intervention: REPLACE-2 randomized trial. JAMA 2003; 289: 853-63.
29. Schiele F, Meneveau N, Gilard M, et al. Intravascular ultrasound-guided balloon angioplasty compared with stent. Immediate and 6-month results of the multicenter, randomized

- Copyright - Il Pensiero Scientifico Editore downloaded by IP 216.73.216.255 Tue, 15 Jul 2025, 23:10:11

Balloon Equivalent to Stent Study (BEST). Circulation 2003; 107: 545-51.
30. Van Gelder IC, Hagens VE, Bosker HA, et al, for the Rate Control versus Electrical Cardioversion for Persistent Atrial Fibrillation Study Group. A comparison of rate control and rhythm control in patients with recurrent persistent atrial fibrillation. N Engl J Med 2002; 347: 1834-40.
31. Packer M, Califf RM, Konstam MA, et al. Comparison of omapatrilat and enalapril in patients with chronic heart failure. The Omapatrilat Versus Enalapril Randomized Trial of Utility in Reducing Events (OVERTURE). Circulation 2002; 106: 920-6.
32. Ross AM, Molhoek P, Lundergan C, et al, for the HART II Investigators. Randomized comparison of enoxaparin, a low-molecular-weight heparin, with unfractionated heparin adjunctive to recombinant tissue plasminogen activator thrombolysis and aspirin: second trial of Heparin and Aspirin Reperfusion Therapy (HART II). Circulation 2001; 104: 648-52.
33. The European Agency for the Evaluation of Medicinal Products (EMEA). Points to consider on switching between superiority and non-inferiority. London, July 27, 2000; CPMP/EWP/482/99.
34. Garattini S, Bertele V, Li Bassi L. How can research ethics committees protect patients better? BMJ 2003; 326: 1199-201.