

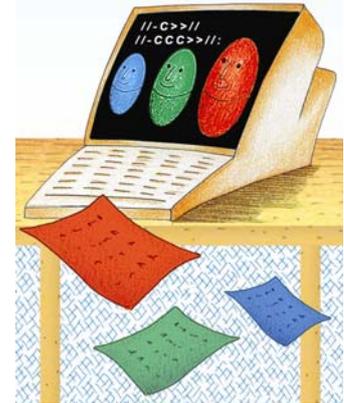
# Dimmi cosa vuoi testare... e ti dirò che studio ti serve

## Studi di superiorità, equivalenza, non inferiorità

Nel confrontare l'efficacia di 2 trattamenti o 2 farmaci, ad es. uno nuovo, che definiamo **intervento (I)** ed uno di riferimento, che definiamo **controllo (C)**, si può ipotizzare che:

- I sia migliore di C
- I sia quasi uguale a C
- I non sia peggio di C

Prima di affrontare in dettaglio come confrontare 2 trattamenti, si farà qualche accenno agli intervalli di confidenza e al "valore p" statistico. **È importante sottolineare che tutti gli elementi di statistica sono presentati in modo intuitivo, senza alcuna pretesa di essere esaustivi.**



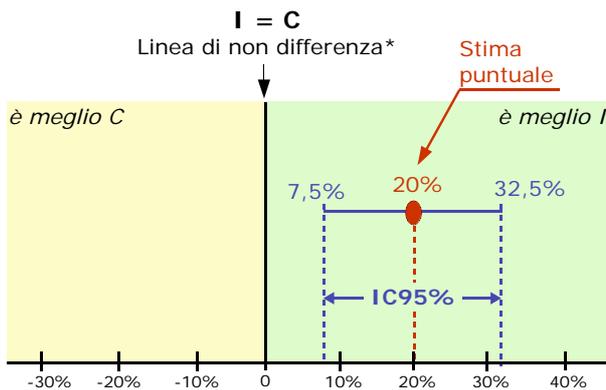
### Intervallo di confidenza: più informativo del "valore p"

In generale, il risultato di uno studio (vedi tabella 1) è solitamente rappresentato da:

- una stima puntuale espressa in termini assoluti o relativi (nell'es.: 20% e 33%);
- un **intervallo di confidenza - IC95%** (nell'es.: da 7,5% a 32,5%);
- un livello di significatività statistica solitamente espresso dal **valore p** (nell'es.: 0,001).

**Tabella 1.** Esempio del risultato di uno studio: sono indicati il valore del risultato, l'IC95% e il valore p.

	pz guariti/tot pz (% pz guariti)	Diff. assoluta %	IC95% della diff. % assoluta	Diff. relativa %	p
intervento (I)	80/100 (80%)	20%	da 7,5% a 32,5%	33%	0,001
controllo (C)	60/100 (60%)				



**Figura 1.** Rappresentazione grafica di un risultato (espresso come differenza assoluta %) e del suo IC95%.

### "Valore p": nessuna informazione sulla rilevanza clinica del risultato

Il "valore p", forse il più noto tra gli indicatori statistici, esprime la probabilità che il risultato ottenuto sia dovuto al caso.

**Per convenzione, si definisce significativo dal punto di vista statistico un risultato accompagnato da un valore di  $p \leq 0,05$  che equivale a dire che vi è al massimo un 5% di probabilità che il risultato osservato sia dovuto al caso.**

Il "valore p" **non fornisce** tuttavia **informazioni** sull'entità dell'effetto osservato e **sulla rilevanza clinica del risultato**: piccole differenze tra trattamenti possono risultare statisticamente significative se studiate in campioni molto grandi.

#### Bibliografia

- Altman DG et al. Statistica medica. Intervalli di confidenza nella ricerca biomedica. Ed. Minerva Medica Torino 2004

### Intervallo di confidenza (IC95%): informazioni importanti sulla rilevanza clinica

Testare un'ipotesi sull'intera popolazione dei pazienti è praticamente impossibile.

Ogni studio, quindi, deve essere fatto su un **campione** dell'intera popolazione e il risultato che si ottiene è una **stima** del valore "reale" sull'intera popolazione di riferimento.

Se si ripetesse 100 volte lo stesso esperimento su 100 campioni diversi estratti dall'intera popolazione si otterrebbero 100 stime diverse.

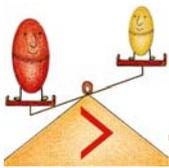
L'**intervallo di confidenza al 95% (IC95%)** esprime l'intervallo di valori entro il quale cadrebbero i risultati 95 volte su 100 e indica approssimativamente la probabilità che il valore "vero" della popolazione di riferimento cada al suo interno.

Stime puntuali e intervalli di confidenza vengono rappresentati graficamente come nella fig.1, collocati in un diagramma ove la linea verticale indica l'assenza di differenza fra i due trattamenti ( $I = C$ ).\*

Se l'intervallo di confidenza è ampio, significa che la stima ottenuta è poco precisa, perché potrebbe anche essere lontana dal valore "vero" che si otterrebbe studiando l'intera popolazione di pazienti.

Uno dei parametri che influenzano maggiormente l'ampiezza dell'IC95% è la numerosità del campione: più il campione è numeroso, più l'intervallo sarà piccolo, più la stima sarà precisa.

**Nota.** \*Quando si confrontano un intervento (I) e un controllo (C), il risultato del confronto tra i due gruppi può essere espresso come **differenza** di frequenze di eventi o come **rapporto** tra le frequenze. Se I e C risultano ugualmente efficaci, nel primo caso il risultato del confronto sarà ( $I-C=0$ ), nel secondo ( $I/C=1$ ).



## Il nuovo è meglio del vecchio? Studi di superiorità

Si ricorre ad uno studio di superiorità quando gli sperimentatori presumono che il nuovo trattamento sia **più efficace** di quello di riferimento; esempio tipico è quello degli **studi contro placebo**.

Se **I** è il **nuovo trattamento** e **C** quello di riferimento, sulla base della rappresentazione grafica della **figura 2**, si potranno avere diversi risultati.

Solo se l'intervallo di confidenza **NON** supera la linea di NON differenza ( $I = C$ ) si potrà affermare che la differenza tra i due trattamenti è significativa dal punto di vista statistico.

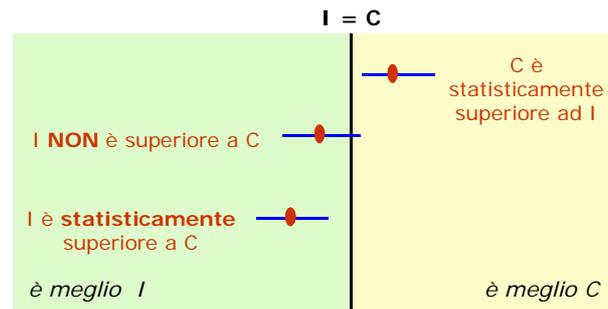


Figura 2. Studi di superiorità.

### Esempio 1. Aspirina e prevenzione primaria dell'infarto miocardico

Il Physicians' Health Study ha valutato l'efficacia nella prevenzione primaria dell'infarto miocardico di aspirina (325 mg/die) verso placebo in 22071 medici americani di sesso maschile. Dopo circa 5 anni di trattamento, i risultati per l'infarto miocardico (fatale e non) sono quelli riportati in **tabella 2b** e **figura 2b**: l'aspirina al dosaggio di 325 mg/die si è dimostrata superiore al placebo nella prevenzione dell'infarto miocardico.

**Tabella 2b.** Physicians' Health Study: sono indicati la stima puntuale del risultato, l'IC95% ed il valore p.

	n.infarti totali/tot pz (%)	Riduzione assoluta del rischio %	IC95%	p
aspirina 325 mg/die ( <b>I</b> )	139/11037 (1,3%)	<b>-0,9%</b>	da -0,6% a -1,2%	0,0000001
placebo ( <b>C</b> )	239/11034 (2,2%)	ovvero: (1,3% - 2,2%)		

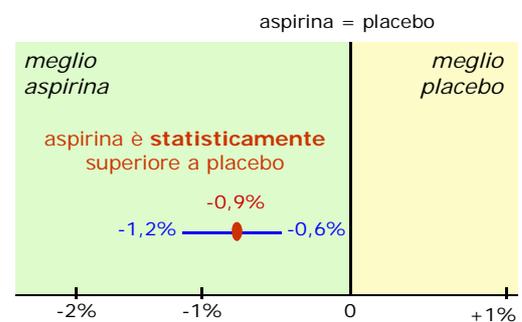


Figura 2b. Rischio di infarto miocardico: aspirina verso placebo.

### Esempio 2. Pregabalin verso placebo nel dolore da neuropatia diabetica

Lo studio di Rosenstock et al. ha valutato l'efficacia di pregabalin (300 mg/die per 8 settimane) verso placebo nel controllo del dolore in 144 pazienti affetti da neuropatia diabetica periferica. L'entità del dolore veniva misurata con una **scala numerica non validata** in cui 0 corrispondeva all'assenza di dolore e 10 al massimo dolore possibile (esito primario).

I risultati dello studio sono presentati in **tabella 2a** e in **figura 2a**: l'intervento (pregabalin 300 mg/die) riduce il dolore rispetto al placebo di circa 1,5 punti. Ma quale è il significato clinico di questa riduzione? (vedi box in fondo).

**Tabella 2a.** Studio di Rosenstock et al.: sono indicati la stima puntuale del risultato, l'IC95% e il valore p.

	punteggio medio	Riduzione assoluta del dolore	IC95%	p
pregabalin 300 mg/die ( <b>I</b> )	3,99 punti	<b>-1,47 punti</b>	da -2,19 punti a -0,75 punti	0,0001
placebo ( <b>C</b> )	5,46 punti	ovvero: (3,99-5,46)		

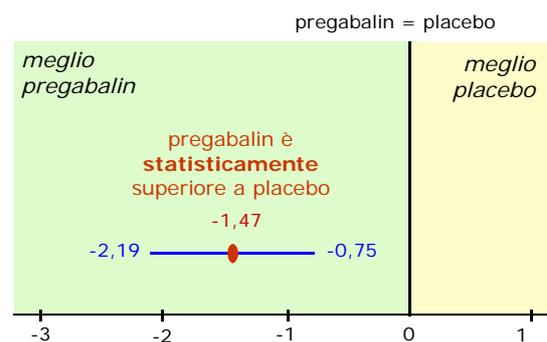


Figura 2a. Differenza nel punteggio medio del dolore: pregabalin verso placebo.



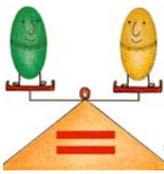
#### Ciò che conta è il significato clinico del risultato

Entrambi gli esempi mostrano risultati statisticamente significativi (infatti il valore p è molto inferiore a 0,05). Nel caso del pregabalin è osservabile una differenza statisticamente significativa con un piccolo campione; nel caso dell'aspirina invece, per raggiungere la significatività statistica è stato necessario coinvolgere un numero molto elevato di soggetti. Tuttavia, **nella pratica non basta la significatività statistica; ciò che conta è il significato clinico del risultato.**

Una diminuzione di quasi l'1% di infarti miocardici è un risultato di interesse clinico, mentre è stato dimostrato che una differenza minore di 2 punti su una scala del dolore ha un significato clinico incerto.

#### Bibliografia

- Rosenstock J. et al. Pregabalin for the treatment of painful diabetic peripheral neuropathy: a double-blind, placebo-controlled trial. *Pain* 2004; 110: 628-638
- Steering committee of the physician's health study research group. Final report on the aspirin component of the ongoing physicians's health study. *N Engl J Med* 1989; 321: 129-35



## Il nuovo è equivalente al vecchio? Studi di equivalenza

Lo scopo è quello di dimostrare che l'intervento (I) è efficace tanto quanto il controllo (C). Poiché dimostrare che due interventi sono esattamente sovrapponibili è impossibile, l'equivalenza deve essere stabilita individuando un intervallo entro il quale devono cadere il risultato e il relativo IC95% (vedi figura 3). Tale intervallo, definito **marginale di equivalenza**, dovrebbe essere piccolo e identificare una differenza non rilevante dal punto di vista clinico. Un margine di equivalenza ad es. del  $\pm 5\%$  significa che la differenza di efficacia (ed il suo IC95%) tra I e C deve essere compresa tra  $-5\%$  e  $+5\%$ : si potrà quindi affermare che I è equivalente a C solo se la stima e il suo intervallo di confidenza pur superando la linea di NON differenza ( $I = C$ ) rimangono all'interno del margine di equivalenza.

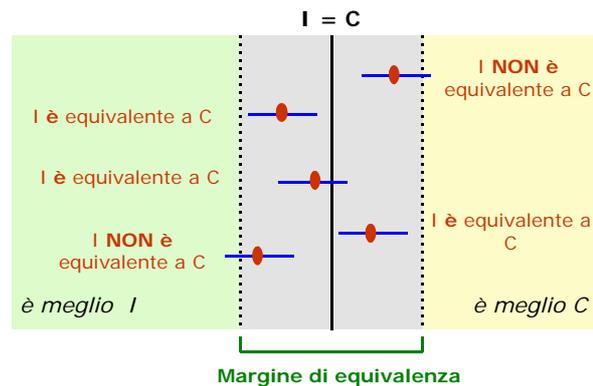


Figura 3. Studi di equivalenza.

### Esempio. Tobramicina e ceftazidima ev: una somministrazione/die equivale a tre?

Lo studio TOPIC si è proposto di dimostrare l'equivalenza fra due diverse modalità di somministrazione di tobramicina + ceftazidima (la stessa dose in monosomministrazione rispetto a 3 somministrazioni giornaliere) nel migliorare il FEV<sub>1</sub>. Lo studio è stato condotto su 219 pazienti con fibrosi cistica e infezione da *Ps. aeruginosa*. Entrambi gli schemi terapeutici sono stati somministrati per 14 giorni. Il risultato per l'esito primario (incremento % del FEV<sub>1</sub> rispetto al basale) è presentato in tabella 3a e in figura 3a.

Tabella 3a. Studio TOPIC: sono indicati il valore del risultato e l'IC95%.

	incremento medio di FEV <sub>1</sub>	Diff. assoluta %	IC95%
1 somm./die (I)	10,4%	0,4% ovvero: (10,4%-10,0%)	da -3,3% a 4,1%
3 somm./die (C)	10,0%		

#### Nota alla tabella

La tabella 3a non riporta il "valore p", ma può essere calcolato anche per gli studi di equivalenza: in questo caso esso esprime la probabilità che i 2 trattamenti differiscano per un valore superiore al margine di equivalenza.

#### Qualche commento...

Il margine di equivalenza (differenza di incremento di FEV<sub>1</sub>) era stato preventivamente fissato nel  $\pm 4\%$ : se il risultato (e il suo IC95%) fosse stato compreso tra  $-4\%$  e  $+4\%$ , i due schemi sarebbero stati da considerare equivalenti. *Nell'esempio, l'estremo superiore dell'IC95% è 4,1%. Pertanto, in termini strettamente statistici, i due schemi non sono da considerare equivalenti.* In realtà, ritenendo trascurabile la differenza tra 4,0% e 4,1%, gli autori concludono per l'equivalenza dei 2 trattamenti.

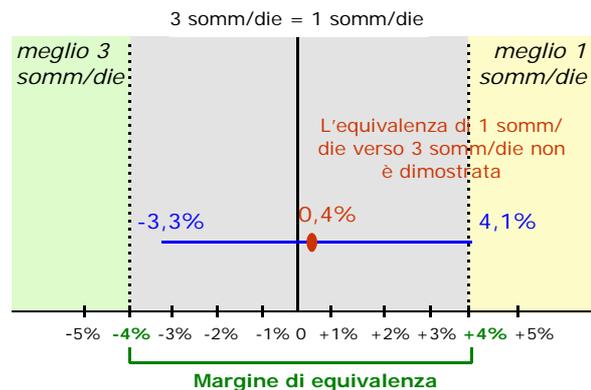


Figura 3a. Differenza di incremento % di FEV<sub>1</sub>.

#### Bibliografia

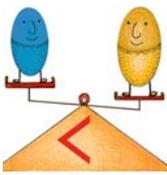
- Greene et al. Claims of equivalence in medical research: are they supported by evidence? *Ann Intern Med* 2000; 132: 715-722
- Piaggio G et al. Reporting of Noninferiority and Equivalence Randomized Trials. *JAMA* 2006; 295: 1152-60
- Smyth A. et al. Once versus three-times daily regimens of tobramycin treatment for pulmonary exacerbations of cystic fibrosis—the TOPIC study: a randomised controlled trial. *Lancet* 2005; 365: 573-78

### Perché gli studi di equivalenza sono poco frequenti?

Il fatto che il margine di equivalenza debba essere piccolo e non rilevante dal punto di vista clinico comporta la necessità di coinvolgere un numero molto elevato di soggetti. Tali studi quindi non sono vantaggiosi in termini di impegno organizzativo e di costi.

Un particolare tipo di studio di equivalenza, ovvero quello di non inferiorità (vedi pagina seguente), richiede campioni più piccoli ed è invece più frequente.





## Il nuovo non è peggio del vecchio? Studi di non inferiorità

Per rispondere al quesito si può utilizzare un particolare studio di equivalenza detto di "non inferiorità". L'obiettivo è quello di dimostrare che l'intervento (I) non è inferiore al controllo (C) accettando quindi che il nuovo farmaco possa essere meno efficace di quello di riferimento.

Meno efficace di quanto? Di un valore predefinito detto **margin** di non inferiorità oltre il quale non devono cadere il risultato ed il suo IC95%. Per definirlo si deve tener conto di diversi elementi quali, ad esempio, il tipo di esito valutato e la sua gravità, l'efficacia degli altri trattamenti disponibili, ecc. E' comunque fondamentale che il margine di non inferiorità identifichi una differenza **non rilevante dal punto di vista clinico**. Ad es. se il margine di non inferiorità è del -5%, si potrà affermare che il risultato è statisticamente significativo (e quindi che la non inferiorità è dimostrata) se la stima puntuale e il suo intervallo di confidenza non cadono oltre il -5%.

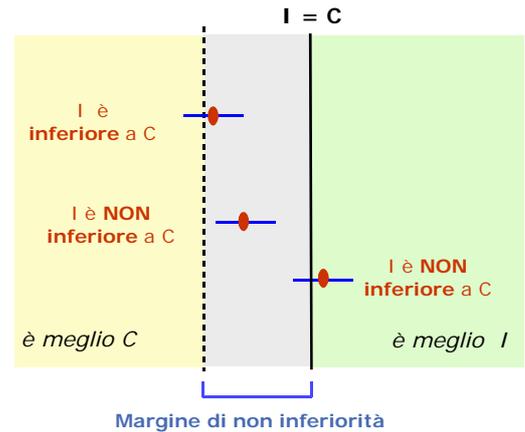


Figura 4. Studi di non inferiorità.

### Esempio. Tigeciclina verso vancomicina+aztreonam nelle infezioni complicate della cute

Lo studio di Breedt et al ha valutato l'efficacia di tigeciclina ev (50 mg x 2/die per 14 giorni) verso vancomicina/aztreonam ev (1 g + 2 g x 2/die per 14 giorni) per il trattamento di infezioni complicate della cute e degli annessi cutanei.

Per definire tigeciclina non inferiore a vancomicina/aztreonam gli sperimentatori avevano stabilito che la differenza assoluta (e il suo IC95%) nella percentuale di risposta clinica tra i due trattamenti non dovesse essere inferiore a -15% (margine di non inferiorità).

I principali risultati dello studio sono illustrati in tabella 4a e figura 4a.

**Tabella 4a.** Risultati dello studio di Breedt et al.: sono riportati la stima puntuale del risultato e l'IC95%.

	pz guariti/ pz trattati (%)	Diff. Assoluta (%)	IC95%
tigeciclina ev (50mg x 2/die) (I)	200/223 (89,7%)	-4,7%	da -10,2% a +0,8%
vancomicina/ aztreonam ev (1g+2g x 2/die) (C)	201/213 (94,4%)	ovvero: (89,7%- 94,4%)	

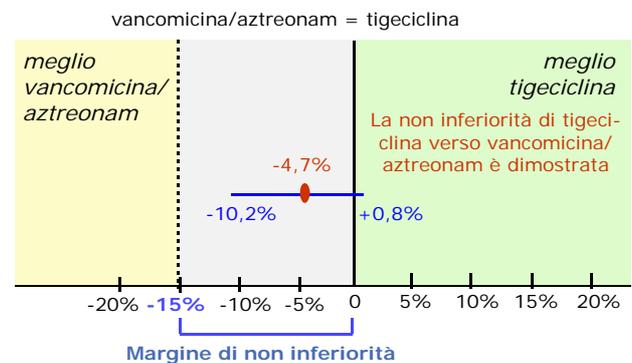


Figura 4a. Differenza nella risposta clinica: tigeciclina vs vancomicina/aztreonam.

#### Qualche commento...

Nello studio illustrato, la tigeciclina si dimostra non inferiore al trattamento di riferimento: infatti pur risultando meno efficace dell'associazione vancomicina/aztreonam, il risultato ed il suo IC95% ricadono all'interno del margine di non inferiorità prestabilito dagli sperimentatori (-15%).

C'è da interrogarsi se sia accettabile considerare clinicamente non rilevante una potenziale minore efficacia del 10,2% della tigeciclina (margine inferiore dell'intervallo di confidenza) rispetto al trattamento di riferimento.

#### Bibliografia

- Piaggio G et al. Reporting of noninferiority and equivalence Randomized Trials. *JAMA* 2006; 295:1152-60
- Sanjay K, Diamone GA. Good enough: a primer on the analysis and interpretation of noninferiority trials. *Ann Intern Med* 2006;145:62-69
- Irving K et al. Design issues in noninferiority/equivalence trials. *Drug Inf J* 1999; 33:1205-18
- Breedt J. et al. Safety and efficacy of tigecycline in treatment of skin and skin structure infections: results of a double-blind phase 3 comparison study with vancomycin-aztreonam. *Antimicrob. agents chemother.* 2005; 49:4658-66
- Garattini S, Bertelé V. Non-inferiority trials are unethical because they disregard patients' interests. *Lancet* 2007; 370:1875-77

### Quando è accettabile uno studio di non inferiorità?



I **presupposti** sono:

- il controllo deve essersi dimostrato sicuro ed efficace rispetto al placebo in condizioni sperimentali analoghe (per tipologia di pazienti, dosaggi utilizzati, ecc.);
- la minore efficacia accettata per il nuovo farmaco/trattamento deve rimanere entro limiti clinicamente non rilevanti;
- l'eventuale minore efficacia del nuovo trattamento deve essere compensata da vantaggi quali maggiore sicurezza, praticità d'uso (ad es. unica somministrazione giornaliera verso più somministrazioni), minor costo, ecc.

#### Pacchetti Informativi sui Farmaci n. 1 / 2008

Periodico di Informazione medica dell'Azienda USL di Modena  
CeVEAS - viale Muratori, 201 - 41100 Modena - Tel 059 435200

Direttore responsabile: Nicola Magrini

Vicedirettori: Anna Maria Marata, Giulio Formoso

Editing e grafica: Barbara Paltrinieri

Elaborazione dati: Claudio Voci

Disegni: Mitra Divshali



Stampa: Premiata Stabilimento Tipografico dei Comuni - Santa Sofia (Fo)  
Chiuso in redazione il 7/1/2008. Tiratura 11.000 copie. Inviato ai medici di medicina generale. Registrazione al Tribunale di Modena n.1787 del 27/02/2006

Questa pubblicazione va citata come:

**Maltoni S, Capelli O, Daya L, Formoso G, Bonacini I, Voci C, Magrini N, Marata AM.** Dimmi cosa vuoi testare... e ti dirò che studio ti serve. Studi di superiorità, equivalenza, non inferiorità. *Pacchetti Informativi sui Farmaci* 2008; 1: 1-4