- Copyright - Il Pensiero Scientifico Editore downloaded by IP 18.224.149.242 Thu, 25 Apr 2024, 18:18:28

Gli studi di non inferiorità

Enzo Ballatori
Docente di Statistica Medica,
Facoltà di Medicina e Chirurgia,
Università di L'Aquila

Negli ultimi anni, gli studi di non inferiorità sono stati al centro di forti critiche sospinte soprattutto dall'abuso che se ne è fatto, sia in termini di inappropriatezza del disegno per lo specifico problema sia per la inadeguatezza dei report degli studi clinici che li hanno adottati, sia per la forzatura degli elementi di soggettività che li caratterizzano, sia per possibili violazioni degli aspetti etici. Insieme ai più rari studi di equivalenza, quelli di non inferiorità costituiscono lo strumento con cui è stata – ed è – veicolata una discreta parte della patologia della ricerca clinica.

Pertanto, nel presente numero della rivista, abbiamo ritenuto opportuno fondere le due rubriche "Casi clinici" e "Statistica per concetti" non solo perché in questo caso i due titoli collimano, ma anche per avere più spazio per la discussione di questi tipi di studi, così frequenti in Oncologia, anche nel campo delle terapie di supporto.

Quando la Statistica perseguiva solo gli scopi della ricerca scientifica, l'ipotesi nulla traduceva simbolicamente l'ipotesi di uguale efficacia dei trattamenti. Quindi la significatività del test statistico indicava che la variabilità osservata (empirica) era così marcatamente superiore a quella teorica da far apparire come assai poco probabile che tale eccesso fosse dovuto unicamente al caso, cioè al campionamento; il che autorizzava a ritenere diversamente efficaci i trattamenti. Viceversa, l'accettazione dell'ipotesi nulla andava interpretata nel senso che non necessariamente i trattamenti a confronto erano ugualmente efficaci, ma semplicemente non vi erano sufficienti ragioni per ritenerli diversamente efficaci. Tale ragionamento ha ampliato, al campo dei fenomeni che presentano variabilità nelle loro manifestazioni individuali, il principio di confutazione di Popper: se i risultati di un esperimento contraddicono quanto previsto da una legge scientifica, questa viene confutata, perdendo così la sua validità generale, mentre risultati in accordo con essa semplicemente la rafforzano, ma non provano che sia vera in quanto si può sempre pensare all'esistenza di un esperimento, non ancora eseguito, che porterà alla sua confutazione.

Successivamente la Statistica venne impiegata come strumento decisionale: fu introdotta la direzionalità dell'ipotesi alternativa, fu sviluppato il concetto di potenza e fu ampliato l'uso dell'ipotesi nulla da puntuale a composta, introducendo così anche i test di equivalenza e di non inferiorità.

Lo studio di superiorità è in linea con la logica suddetta. In esso, l'ipotesi nulla, traducendo l'ipotesi di uguale efficacia dei trattamenti, è un'ipotesi semplice: i parametri delle popolazioni target (quelle cui vanno riferiti i risultati, cioè le popolazioni di pazienti presenti e futuri, che ricevono l'uno o l'altro dei due trattamenti a confronto) sono posti uguali tra loro. Esemplificando: se i trattamenti sono ugualmente efficaci, c'è da attendersi che, nelle popolazioni target, la percentuale di successi terapeutici sia la stessa.

Invece, si definisce l'equivalenza introducendo un intervallo simmetrico intorno allo zero, detto margine di equivalenza, da $-\Delta$ a $+\Delta$, entro cui deve cadere la differenza tra i parametri affinché i trattamenti possano ritenersi equivalenti. Operativamente, i due trattamenti si definiscono equivalenti se l'intervallo di confidenza della differente efficacia osservata ricade per intero nel margine di equivalenza, ossia è compreso tra $-\Delta$ e $+\Delta$.

Ad esempio, se la probabilità di proteggere dalla nausea ritardata fosse del 60% per la terapia antiemetica standard, si potrebbe definire un margine di equivalenza del 10%, così da considerare equivalente al trattamento standard una nuova profilassi antiemetica che ottenesse una percentuale di protezione dal 50 al 70%. Se l'intervallo di confidenza (con coefficiente $1-\alpha$, generalmente fissato al 95%) della differenza tra la protezione fornita al gruppo di trattamento e quella relativa al gruppo di controllo ricadesse nell'intervallo [–10%, +10%], l'equivalenza sarebbe provata.

Gli studi di non inferiorità costituiscono un caso particolare degli studi di equivalenza, in cui l'estremo superiore del margine di equivalenza non è più prefissato, ossia viene considerato l'intervallo $(-\Delta, +\infty)$, dove $-\Delta$ è detto margine di non inferiorità. Riprendendo l'esempio, la nuova profilassi antiemetica è detta non inferiore a quella standard se l'intervallo di confidenza (con coefficiente $1-2\alpha$, cioè del 90%, perché l'estremo superiore non ha più interesse) della differenza tra la percentuale di protezione del nuovo trattamento e quella relativa al trattamento standard non include il valore di –10%, altrimenti la non inferiorità non è dimostrata.

In parole semplici, il nuovo trattamento è definito come "equivalente" rispetto al trattamento standard, anche se ha un'efficacia di "poco" superiore o inferiore; è definito, invece, come "non inferiore" anche se ha un'efficacia di "poco" inferiore a quella della terapia standard: l'essenziale è che la sua minore efficacia non ecceda quella prestabilita (cioè $-\Delta$), altrimenti la sua non inferiorità non è provata.

A parte un personale prurito epistemologico a parlarne,

- Copyright - Il Pensiero Scientifico Editore downloaded by IP 18.224.149.242 Thu, 25 Apr 2024, 18:18:28

gli studi di equivalenza e di non inferiorità presentano limitazioni – sottolineate da più autori – che richiedono adeguate contromisure.

a) Conseguenze della scelta di Δ

Consideriamo gli studi di non inferiorità, avvertendo che gli studi di equivalenza, pur con un diverso formalismo, conservano la stessa logica.

Sia C una misura di efficacia (ad es., la percentuale di successi terapeutici) calcolata nella popolazione (target) di pazienti presenti e futuri, nelle condizioni di interesse, sottoposti al trattamento di controllo (la migliore terapia esistente) e sia T la stessa misura calcolata nella popolazione target dei pazienti che ricevono il nuovo trattamento.

L'ipotesi nulla è

H_0: $C - T \geq \Delta$,

mentre quella alternativa è

H_1: $C - T < \Delta$,

così che respingendo l'ipotesi nulla si perviene ad un verdetto di non inferiorità.

Quindi, negli studi di non inferiorità, l'ipotesi nulla è che l'efficacia del trattamento di controllo sopravanzi quella della nuova terapia di una quantità non inferiore a Δ .

Da come sono formulate le suddette ipotesi, si evince che Δ esprime il concetto di "differenza minima clinicamente rilevante" che è sempre alla base del calcolo della dimensione del campione, anche negli studi di superiorità. Nel contesto degli studi di equivalenza e di non inferiorità, però, si aggiunge un significato ancor più importante perché il margine Δ è connaturato a ciò che deve essere provato. Così, negli studi di superiorità, un valore esagerato di Δ conduce ad una riduzione di potenza dello studio (cioè della possibilità di accorgersi della reale differenza di efficacia dei due trattamenti), mentre negli studi di equivalenza o di non inferiorità fa correre un ulteriore rischio: quello di dichiarare "equivalente" o "non inferiore" il nuovo trattamento anche quando esso è notevolmente meno efficace del trattamento standard.

L'ipotesi nulla è sempre quella che il ricercatore desidera falsificare. È allora evidente che volendo provare la non inferiorità del nuovo trattamento rispetto allo standard, possono essere adottate numerose strategie che tendano ad attutire la maggiore efficacia del trattamento di controllo. La più immediata tra queste è fissare un esagerato valore di Δ. In tal modo, diventa più facile respingere l'ipotesi nulla perché non solo costituisce un ostacolo più difficile da superare per la differenza tra C e T, ma richiede anche una dimensione campionaria ridotta che contribuisce ad impedire di scoprire la reale differente efficacia dei due trattamenti.

La scelta di Δ è particolarmente delicata perché risente di un'inevitabile componente di soggettività e va discussa nel protocollo. Non solo dovrebbe essere supportata da considerazioni cliniche condivise dalla comunità scientifica, ma andrebbe anche ascoltato il paziente per indagare, magari con tecniche di tipo *standard gamble*, quale sia la riduzione di efficacia, rispetto alla terapia standard, che egli considera accettabile per assicurarsi gli altri maggiori benefici del nuovo trattamento.

b) Provata efficacia del trattamento standard

Un divieto assoluto all'uso degli studi di equivalenza o di non inferiorità si ha quando l'efficacia del trattamento standard non è sufficientemente provata. In tal caso, infatti, si corre il rischio di dichiarare il nuovo trattamento "equivalente" (o "non inferiore") allo standard senza sottolineare che anch'esso potrebbe essere inefficace o poco efficace: lo studio diventa inutile ed i suoi risultati forse dannosi.

Come negli studi di superiorità, in quelli di equivalenza o di non inferiorità, è particolarmente delicata la scelta del *comparator*, in quanto se l'efficacia del nuovo trattamento venisse confrontata con quella di un trattamento sub-ottimale, l'attestata equivalenza (non inferiorità) potrebbe comportare una inferiorità rispetto al trattamento standard. I risultati dello studio, dunque, non aggiungerebbero alcuna conoscenza utile per la pratica clinica.

c) Interpretazione dei risultati

Spesso la riduzione di efficacia prevista dal trattamento standard può essere più espressiva se tradotta in termini relativi. Nell'esempio sull'incidenza della nausea ritardata esposto sopra, il trattamento standard garantiva una protezione del 60%, mentre il margine di non inferiorità era fissato a –10%, considerando così non inferiore il nuovo trattamento se riusciva a proteggere almeno il 50% dei pazienti. In termini relativi, tale assunzione sembra accettabile: per essere dichiarato non inferiore, il nuovo trattamento può avere un'efficacia ridotta di circa il 16% (= 10/60) rispetto allo standard.

Ora immaginiamo che l'effetto placebo sia importante (e questo è un caso frequente nelle terapie di supporto) e pari al 30% per entrambi i trattamenti. In tal caso l'efficacia "farmacologica" del trattamento standard risulta del 30% (60 – 30), mentre quella del nuovo trattamento del 20% (50 – 30), nella peggiore delle ipotesi. Quindi la reale diminuzione relativa di efficacia non è più del 16%, ma si raddoppia, diventando del 33% ([30 – 20]/30 = 10/30). C'è da chiedersi se questa reale riduzione di efficacia consenta ancora di indicare il nuovo trattamento come non inferiore rispetto allo standard.

d) Criteri di valutazione dei pazienti

Negli studi di superiorità, il più adottato criterio di valutazione dei pazienti è quello per intenzione a trattare (Intention To Treat, ITT): vanno valutati tutti i pazienti una volta randomizzati, indipendentemente, quindi, dal fatto che rispettassero o meno i criteri di selezione e dalla loro *compliance* al trattamento cui erano stati assegnati. Tale criterio, se da un lato consente di evitare distorsioni da selezione e di fornire risultati assai prossimi a quelli che si otterranno successivamente nella pratica clinica, dall'altro introduce un rumore che potrebbe celare il vero valore del tratta-

- Copyright - Il Pensiero Scientifico Editore downloaded by IP 18.224.149.242 Thu, 25 Apr 2024, 18:18:28

mento. Ad esempio, con il criterio ITT sono valutati anche i pazienti che hanno assunto il farmaco solo una o poche volte e poi hanno cessato di assumerlo per svariate ragioni, tra cui la possibile presenza di eventi avversi. Negli studi di superiorità il rumore così prodotto è da riguardarsi come ulteriore elemento di prudenza nella valutazione comparativa, ma negli studi di equivalenza o di non inferiorità gioca, al contrario, a favore del nuovo trattamento, perché rischia di nascondere reali differenze di efficacia.

Negli studi di equivalenza o di non inferiorità è dunque necessario procedere anche con la valutazione *per protocol*, che considera valutabili tutti i pazienti che hanno completato lo studio e sono stati trattati secondo quanto previsto dal protocollo.

In conclusione, negli studi di equivalenza o di non inferiorità i pazienti andrebbero valutati secondo entrambi i criteri ed il nuovo trattamento andrebbe dichiarato equivalente o non inferiore allo standard solo se il verdetto ottenuto con la valutazione ITT concordasse con quello derivante dalla valutazione *per protocol*.

Per inciso, anche una cattiva conduzione dello studio, che provocasse scarsa aderenza ai trattamenti, elevato numero di drop out, arruolamento prevalente di pazienti che difficilmente risponderanno, imprecisione nella registrazione degli eventi, ecc., produce rumore che è funzionale alla prova di equivalenza o di non inferiorità.

e) Aspetti etici

Studi di equivalenza o di non inferiorità non dovrebbero mai essere condotti nel caso di malattie gravi, in cui la possibile diminuita efficacia del nuovo trattamento si potrebbe tradurre in morti aggiuntive.

Spesso si conosce poco del nuovo trattamento, in termini di efficacia e sicurezza. Il paziente cui venisse somministrato correrebbe il duplice rischio di assumere un farmaco che non solo è verosimilmente un po' meno efficace della terapia standard, ma che è anche meno tollerabile.

Nel contesto di tali studi, l'ottenimento del consenso informato è dunque un momento assai delicato, in quanto occorre chiarire al paziente anzitutto cosa sia uno studio di equivalenza o di non inferiorità, e, soprattutto, che gli potrà essere somministrato un trattamento verosimilmente "un po' " meno efficace della terapia standard (che, data la sua condizione, avrebbe comunque ricevuto) e sulla cui sicurezza si sa ancora poco. In tali condizioni ci si può chiedere quale paziente possa esprimere il proprio consenso a partecipare e, per quelli che aderiscono, quali siano le motivazioni della loro decisione. Per quanto concerne le motivazioni, andrebbe esclusa quella eticamente più rilevante: il benessere dei pazienti futuri; infatti, i risultati di uno studio di equivalenza o di non inferiorità non consentono di individuare trattamenti migliori dello standard, ma solo terapie che sono di efficacia equivalente o non inferiore a quelle già esistenti. Resta così a motivare la scelta del paziente partecipante, soprattutto la fiducia che egli ha nei confronti del medico, fiducia che non deve essere in alcun modo tradita: è quindi necessario che non solo i Comitati Etici costituiscano un filtro di giusta severità nei confronti di tali studi, ma anche che il medico sperimentatore valuti con estrema attenzione la sua partecipazione allo studio alla luce del reale interesse dei propri pazienti.

f) Comunicazione dei risultati

"Equivalenza" è sinonimo di uguale efficacia, ma, come si è visto, in questo contesto non è così. Non inferiorità fa pensare ad un trattamento che ha efficacia superiore o, tutt'al più uguale alla terapia standard, ma non è così. L'uso di questi termini ingenera anzitutto il rischio di veicolare una falsa informazione a vantaggio del *marketing* dell'industria. Inoltre, una volta dichiarato non inferiore il nuovo trattamento potrebbe fungere da *comparator* in un successivo studio. Se ciò si ripetesse, dopo alcuni passaggi si potrebbe giungere a sperimentare contro placebo.

Esistono studi di non inferiorità che costituiscono una buona fisiologia della ricerca clinica, ma la maggior parte di essi è certamente suscettibile di critiche anche sostanziali.

Vi sono alcuni argomenti che possono essere legittimamente indagati con studi di non inferiorità, quando nelle fasi precedenti, il nuovo trattamento, rispetto alla terapia standard, ha presentato un'efficacia similare o di poco inferiore, ma con alcuni vantaggi che potrebbero consistere in una migliore tollerabilità del nuovo trattamento, o in diverse e più accettabili modalità e via di somministrazione della terapia che accrescano la *compliance* del paziente, o, ancora, in termini di riduzione dei costi.

Tutti gli altri casi dovrebbero essere valutati con un'attenzione particolare anzitutto dal medico sperimentatore prima di decidere sulla sua partecipazione allo studio, poi dai Comitati Etici.

Malgrado le pesanti limitazioni esposte, alcune delle quali erano già state sottolineate in un'editoriale di 6 anni fa (Gotzsche PC. Lessons from and cautions about non inferiority and equivalence randomized trials. JAMA 2006; 295: 1172-74), c'è da chiedersi perché il peso percentuale degli studi di non inferiorità rispetto al totale della ricerca clinica comparativa sia così cresciuto.

La risposta è duplice: da un lato tali studi sono funzionali agli interessi dell'industria, dall'altro la comunità scientifica e gli organismi deputati a sorvegliare sugli interessi del paziente non hanno ancora assunto i necessari atteggiamenti di giusta severità nel loro controllo.

Ingenti risorse sono state impiegate dall'industria per giungere a studi di fase II i cui risultati, però, si sono rivelati non così promettenti da indurla ad investire in uno studio di superiorità di fase III, che potrebbe negare al nuovo trattamento l'ingresso nella pratica clinica. Poiché le autorità regolatorie lo consentono, l'industria procede con uno studio di non inferiorità (meno costoso) così da poter entrare sul mercato e da acquisire informazioni utili per decidere successivamente sulla programmazione di uno studio di superiorità. Questa, seppure in parte criticabile, è la fisiologia. La patologia nasce quando l'industria è ragionevolmente certa che il nuovo trattamento sia inferiore a quello standard, in

- Copyright - Il Pensiero Scientifico Editore downloaded by IP 18.224.149.242 Thu, 25 Apr 2024, 18:18:28

termini di efficacia e/o di sicurezza, ma ha bisogno di entrare sul mercato per coprire le spese sostenute per lo sviluppo del farmaco. In tali casi lo studio di non inferiorità si presta meglio degli altri a possibili artifici che consentono di raggiungere gli obiettivi prefissati.

Proprio a causa di questa particolare elasticità degli studi di non inferiorità, è necessario che gli altri attori del processo di ricerca clinica, dal medico sperimentatore, ai Comitati Etici, ai *referee* della rivista che pubblicherà i risultati, abbiano come obiettivo primario la protezione del paziente e siano attrezzati per valutare con rigore se i risultati dello studio siano accettabili o meno. Ma che tali meccanismi di protezione del paziente non siano ancora entrati in esecuzione è attestato dal grande numero di studi di non inferiorità pubblicati di recente, malgrado già 6 anni fa una ricerca abbia evidenziato le gravi carenze riscontrate in molti studi di equivalenza e di non inferiorità (Le Henanff A et al. Quality of reporting of noninferiority and equivalence randomized trials. JAMA, 2006; 295: 1147-51).

Alla luce delle limitazioni degli studi di equivalenza o non inferiorità, si può costruire una *check list*, forse non esaustiva, ma certamente utile per valutare protocollo e risultati di tali tipi di studi. Occorre verificare che il protocollo dello studio ed il report che descrive i risultati ottenuti indichino chiaramente:

1. il tipo di studio (di equivalenza, di non inferiorità) e le ragioni di tale scelta;
2. la motivazione della scelta del *comparator*, sia sottolineando che è il miglior trattamento esistente, sia citando le prove della sua efficacia;
3. la definizione del margine di equivalenza o di non inferiorità, esponendo le ragioni della sua scelta, in particolare della sua accettabilità clinica;
4. le procedure che hanno portato al calcolo della dimensione del campione, tenendo conto del margine Δ;
5. i criteri di valutazione dei pazienti, ITT e per *protocol*, e che entrambi verranno (sono stati) usati per la valutazione di equivalenza o non inferiorità.

Per chi valuta i risultati della ricerca (*referee*, membri delle *Consensus Conference* chiamati a definire le linee guida di trattamento, clinici), la prima richiesta da fare, è la disponibilità del protocollo dello studio. In verità, il protocollo è registrato, quindi è di pubblico dominio, solo che è complicato trovarlo. Sarebbe quindi necessario che l'articolo che pubblica i risultati contenga le coordinate per rintracciare agevolmente il protocollo registrato. La comparazione tra quanto riportato nel lavoro e quanto contenuto nel protocollo è importante perché potrebbe verificarsi che:

- i risultati si riferiscano ad uno studio di superiorità, ma che non avendo raggiunto l'obiettivo si sia deciso di pubblicarli come fossero riferiti ad uno studio di non inferiorità. La metodologia usata nei due tipi di studi è diversa, sia in termini di analisi statistica che nel calcolo della dimensione del campione. Programmare uno studio di un certo tipo ed utilizzare i risultati come si trattasse di uno studio di tipo diverso è del tutto inaccettabile;
- l'endpoint principale dello studio sia stato scambiato con uno secondario;
- la soglia Δ adottata nell'esposizione dei risultati non coincida con quella indicata nel protocollo, ma sia stata stabilita *ex post*, su una base di convenienza, alla luce dei risultati ottenuti. In altre parole, potrebbe essere accaduto che per il valore Δ fissato nel protocollo non sia stata raggiunta la prova di non inferiorità, mentre considerando un margine leggermente più ampio, il nuovo trattamento possa essere dichiarato non inferiore alla terapia standard.

Conclusioni

Il lettore si chiederà se quanto ha letto sia terrorismo metodologico o, se non lo fosse, perché si continuino a programmare e, soprattutto, ad accettare studi di equivalenza (pochi) e studi di non inferiorità (molti, troppi). A parte la mia personale idiosincrasia per tali tipi di studi (ogni volta che rifletto sulla formulazione dell'ipotesi nulla esposta al punto (a), mi sembra di trovarmi all'interno di un quadro di Escher), si potrebbe rispondere al secondo quesito con la metafora del bambino e dell'acqua sporca, in quanto esiste una metodologia rigorosa che in alcune – sebbene non frequenti – occasioni potrebbe tornare utile per raggiungere in modo economico determinati obiettivi. Quindi, gettarla via perché molti studi di non inferiorità sono stati pianificati e condotti in modo non corretto, ovvero perché tali studi si prestano particolarmente ad essere strumentalizzati per finalità di mercato, mi sembra davvero uno spreco.

Si tratta quindi di esaminare caso per caso, vigilando che il disegno di equivalenza o di non inferiorità sia realmente appropriato, e verificando il rispetto di tutte le norme che garantiscono scientificità alla ricerca clinica.

Questo, in fondo, è l'atteggiamento che il Comitato Nazionale per la Bioetica ha tenuto nell'esprimere un parere sulla sperimentazione clinica con disegno di non inferiorità, il 24 aprile 2009. La critica, sempre assai rigorosa, è stata così serrata che, in sede di conclusioni, ci si sarebbe atteso l'ostracismo – o almeno un anatema – verso tale tipo di studi; invece, è stata adottata la raccomandazione che "gli studi di non inferiorità siano illustrati con maggiore trasparenza ed analizzati con attenzione da parte dei Comitati Etici, che debbono vigilare in modo particolare affinché gli interessi del paziente non siano subordinati ad altri interessi, inclusi quelli commerciali e dello sponsor".

Mi sembra una posizione decisamente equilibrata e degna di un organismo di così alto prestigio. •