

2

Interpretazione di un rapporto su uno studio clinico

Desmond G. Julian, Stuart J. Pocock

Introduzione

Quando si avvia uno studio clinico, la prima preoccupazione è quella di assicurarsi che il suo disegno e la sua conduzione siano conformi alla migliore pratica corrente; l'analisi e la relazione sui suoi risultati sono tuttavia di pari importanza. Come descritto nel primo capitolo, i principi del disegno di uno studio sono oggi formulati chiaramente; una minore attenzione è stata posta però al modo in cui vengono presentati gli studi e vengono fatte le estrapolazioni. In questo capitolo verranno discussi questi problemi e verranno fornite linee-guida la cui applicazione permetterà di giudicare la qualità di una relazione su uno studio clinico. Nella valutazione dell'utilità di qualsiasi relazione su uno studio clinico, ci sono tre ambiti di problemi:

1. *Validità interna.* Lo studio ha confrontato in modo corretto i trattamenti studiati sui pazienti reclutati? Ossia, il disegno e la conduzione erano affidabili e non distorti, il follow-up e l'analisi erano completi e appropriati, le conclusioni erano coerenti con i risultati?
2. *Dimensioni dello studio.* I pazienti studiati, il follow-up, gli eventi clinici osservati erano sufficienti per determinare un confronto preciso (stima) di qualsiasi differenza esistente tra i trattamenti?
3. *Validità esterna.* I risultati dello studio erano di diretta rilevanza per determinare una strategia di trattamento per i pazienti futuri? Ossia, i pazienti arruolati, i regimi di trattamento, le misure delle risposte dei pazienti erano appropriati per valutare il trattamento come inteso nella pratica clinica di routine?

Validità interna

Randomizzazione

a. *I pazienti sono stati assegnati al trattamento in modo randomizzato?* Questo è, naturalmente, un prerequisito essenziale per uno studio non distorto.

b. *I dettagli forniti sul metodo di randomizzazione sono sufficienti a stabilire con affidabilità che la randomizzazione è stata condotta in modo corretto per tutti i pazienti?* Ci sono numerosi modi in cui un processo di randomizzazione può funzionare male, e quindi la sezione «metodi» di una relazione su uno studio clinico dovrebbe contenere una breve spiegazione di come è stata preparata e utilizzata la randomizzazione.

c. *Il metodo di assegnazione era tale per cui l'allocazione prevista poteva essere conosciuta dal paziente o dallo sperimentatore (il clinico) prima del completamento della randomizzazione?* Bisogna essere particolarmente attenti alla possibilità che il medico che arruola il paziente possa conoscere o prevedere a quale gruppo di trattamento verrà assegnato il paziente successivo, o che possa togliere dallo studio il paziente una volta conosciuta la sua assegnazione. In uno studio in doppio cieco, è improbabile che questo rappresenti un problema, ma in alcuni studi (specialmente quelli che riguardano la chirurgia o l'angioplastica), è necessario l'uso di un metodo assolutamente sicuro di randomizzazione (come ad esempio la randomizzazione telefonica), anche nel caso sorgano difficoltà se un paziente cambia idea dopo avere dato il suo consenso. Questo sottolinea l'importanza del principio dell'intention to treat, anche se, nel caso che questo comportamento diventi frequente, la validità dello studio può essere minacciata.

d. *È stata verificata la comparabilità di base dei gruppi di trattamento?* Questa è un'informazione cruciale e le variabili di base più importanti dovrebbero essere incluse nel rapporto. Negli studi di grandi dimensioni è improbabile che le caratteristiche di base non risultino confrontabili, ma anche in questo caso si dovrebbe presentare l'evidenza che il «giusto comportamento» della randomizzazione ha prodotto gruppi bilanciati di trattamento.

Mascheramento

a. *Il trattamento randomizzato ricevuto da ciascun paziente era mascherato adeguatamente? Per esempio, se appropriato, lo studio era in doppio-cieco?* Quando possibile, gli studi clinici dovrebbero essere condotti in doppio-cieco. Non è sempre possibile, come nel caso di procedure come la chirurgia cardiaca e l'angioplastica. Ma ci sono altre situazioni nelle quali gli sperimentatori considerano il cieco non praticabile, come ad esempio gli studi sugli anticoagulanti. Meno credito può essere dato agli studi clinici non in doppio-cieco, dato che è probabile che i pazienti vengano curati e valutati in modo diverso nei diversi gruppi, per quanto riguarda tutto ciò che è al di fuori dei trattamenti randomizzati. Se l'obiettivo principale è un evento oggettivo, come per esempio la mortalità, il non mascheramento può non essere così importante; grande cautela va invece posta se gli obiettivi prefissati sono soggettivi, come ad esempio la necessità di chirurgia o aspetti della qualità della vita.

b. *Lo staff clinico era al corrente dell'assegnazione del trattamento?* Come già accennato sopra, se lo staff clinico coinvolto è al corrente di, o può sospettare, quale trattamento è stato assegnato, ciò può provocare distorsioni nella valutazione e influenzare il tipo di trattamento concomitante somministrato. Si incontrano difficoltà negli studi ove uno dei trattamenti sperimentali ha effetti collaterali frequenti e caratteristici (come ad esempio il rossore con i farmaci vasodilatatori), o richiede un tipo di monitoraggio specifico e complesso (come con gli anticoagulanti); tentativi di preservare il cieco possono essere in questi casi solo parzialmente utili per ridurre la possibilità di distorsioni.

c. *I pazienti erano al corrente dell'assegnazione del trattamento?* È inevitabile che lo siano negli studi non in cieco. Negli studi in cieco, in genere non lo sono; dato che tuttavia dovrebbero venire informati dei benefici potenziali e degli effetti avversi del trattamento che gli potrebbe venire assegnato, spesso non è difficile per loro indovinare l'assegnazione del trattamento. È più probabile ottenere un confronto più corretto se il paziente non è esplicitamente informato riguardo al trattamento assegnatogli.

d. *Chi ha valutato gli eventi era al corrente dell'assegnazione del trattamento?* Specialmente nel caso in cui siano coinvolti obiettivi clinici soggettivi («soft endpoints»), è altamente desiderabile che coloro che assegnano gli eventi a categorie prespecificate siano all'oscuro del gruppo di trattamento al quale il paziente appartiene. In genere, questo

lo si ottiene utilizzando un «Endpoint Committee» (comitato per gli eventi clinici), che rivede i dati di rilievo senza conoscere l'allocazione del trattamento. Ciò è particolarmente importante quando sono coinvolti eventi come l'infarto miocardico e la morte improvvisa (in opposizione alla non-improvvisa), perché la conoscenza del trattamento ricevuto potrebbe influenzare la scelta della categoria diagnostica.

e. *I dati individuali dei pazienti sono stati inseriti nel database in cieco rispetto al trattamento assegnato (e chi ha condotto l'analisi era all'oscuro del trattamento assegnato)?* È altamente desiderabile che chi inserisce i dati nel database sia all'oscuro del trattamento assegnato ai pazienti. È in corso un dibattito sull'opportunità che lo statistico che analizza i risultati dello studio (ed anche il comitato di monitoraggio dei dati che valuta i risultati ad interim) sia cieco rispetto ai trattamenti, ma al presente la maggior parte degli studi non segue questa via.

Follow-up dei pazienti

a. *Ci sono pazienti randomizzati che non hanno ricevuto il trattamento assegnato?* Succede che alcuni pazienti non ricevano il trattamento al quale sono stati randomizzati nello studio. I motivi comprendono il decesso o altri eventi clinici che si verificano prima che il trattamento possa essere iniziato, il rifiuto del paziente o del medico, gli errori nella somministrazione. La pubblicazione relativa ai risultati dovrebbe includere informazioni su questi pazienti, ed è in genere appropriato includerli nell'analisi condotta secondo il principio di «intention to treat». Tuttavia possono sorgere problemi non risolvibili se ci sono stati errori nel processo di randomizzazione, come ad esempio una mancata assegnazione del codice al paziente o l'assegnazione del trattamento sbagliato, oppure la distruzione della fiala contenente il farmaco prima della somministrazione della dose iniziale. Tali problemi dovrebbero venire dichiarati ed è oggetto di discussione se questi casi vadano inclusi nello studio, o se possano essere utilizzati metodi di compenso.

b. *È specificato con chiarezza quanti pazienti sono stati seguiti, quando sono avvenute le visite di valutazione, e per quale periodo di tempo sono stati raccolti gli eventi clinici?* La durata proposta per lo studio e il tipo di visite previste per ciascun paziente dovrebbero essere state dichiarate nel protocollo; per esempio ciascun paziente verrà studiato per un minimo di x giorni, mesi o anni, e seguito per un periodo

prefissato o fino a una data di calendario, fine dello studio. È importante che il periodo di tempo utilizzato nell'analisi sia specificato con chiarezza e che qualsiasi deviazione dall'intento originale del protocollo venga giustificato. In particolare non si deve cambiare retrospettivamente la durata di uno studio allo scopo di fornire i risultati statisticamente più significativi.

c. *Complessivamente, quanti pazienti non sono stati aderenti al trattamento, sono stati trasferiti ad un altro trattamento, o sottratti al follow-up?* Ciascuno di questi punti dovrebbe essere indicato chiaramente nel rapporto. È importante sottolineare che, in accordo con il principio dell'intention to treat, l'obiettivo è di non sottrarre nessun paziente a un follow-up completo e all'analisi, anche se è stato sottratto al trattamento assegnato. Le perdite al follow-up dovrebbero essere minimizzate, ed anche documentate nel rapporto finale. La non aderenza al trattamento è comune negli studi clinici e può distorcere i risultati; dovrebbero essere instaurati metodi per scoprirla. Negli studi clinici su farmaci, questi possono includere il conteggio delle pillole; esistono tuttavia metodi più sofisticati che dovrebbero venire attuati quando praticabili.

d. *Sono indicate chiaramente le ragioni della non aderenza al trattamento, dei trasferimenti, della sottrazione di pazienti dall'analisi?* In particolare, se gli effetti collaterali del trattamento o la mancanza di miglioramento del paziente sono ragioni che contribuiscono sostanzialmente a queste deviazioni dal protocollo, esse forniscono suggerimenti importanti sugli aspetti meno benefici del trattamento randomizzato, ed è quindi essenziale una documentazione completa di queste conseguenze avverse.

Approcci all'analisi statistica

a. *L'obiettivo primario prespecificato è stato dichiarato ed analizzato chiaramente?* La maggior parte degli studi ha numerose misure di interesse, e ciò può causare problemi di interpretazione se non sono specificate le priorità. In particolare, in assenza di una definizione preventiva delle priorità, uno sperimentatore disonesto potrebbe essere incline a enfatizzare la differenza più significativa tra i trattamenti, tra quelle osservate fra molti obiettivi. Quindi è in genere appropriato predefinire un singolo obiettivo principale e la sua analisi. In alcune circostanze uno studio può avere due (o, raramente, più di due) obiettivi chiave di interesse primario.

b. *È stato applicato il principio dell'intention-to-treat nell'analisi principale?* Analisi «intention-to-treat» significa che tutti i pazienti randomizzati a ciascun gruppo di trattamento (e tutti i loro tempi di follow-up) vengono inclusi nell'analisi comparativa, prescindendo dal fatto che abbiano o meno ricevuto il trattamento assegnato, in modo completo o parziale. Questo principio di analisi è considerato essenziale per ottenere un paragone non distorto tra «strategie» di trattamento, nel modo in cui verranno realmente implementate. Se numeri sostanziali di pazienti deviano dal trattamento assegnato, e soprattutto se ciò è distribuito in modo disuguale tra i trattamenti, analisi supplementari possono risultare utili per esplorare il risultato in relazione alla aderenza al trattamento (per esempio con analisi «on treatment»), ma l'enfasi principale andrà comunque posta sull'analisi «intention-to-treat».

c. *Viene specificata e discussa la dimensione prevista dello studio?* Nel correlare i risultati pubblicati dello studio con il disegno previsto è utile vedere un razionale documentato per il numero prespecificato di pazienti necessario e in che misura questo numero si sovrappone al numero dei pazienti realmente inclusi nello studio. La credibilità dei risultati riportati risulta in genere aumentata se il numero dei pazienti è coerente con le intenzioni motivate scientificamente a priori.

d. *Lo studio è stato concluso più precocemente o più tardivamente rispetto a quanto pianificato?* Esistono valide ragioni per chiudere uno studio più precocemente rispetto a quanto pianificato; per esempio, se i risultati non in cieco dimostrano che il beneficio o il rischio di uno dei trattamenti è così grande che non sarebbe stato etico proseguire. Una chiusura per questi motivi è resa più credibile se è giustificata da linee-guida statistiche predefinite. Una chiusura può essere giustificata da una quantità di altre ragioni, che comprendono una evidenza di beneficio o di rischio emersa al di fuori dello studio, l'incapacità di reclutare pazienti nello studio, considerazioni di tipo finanziario. È tuttavia inaccettabile che una casa farmaceutica chiuda uno studio perché ha «perso interesse».

e. *Perché lo studio viene riferito ora?* Esistono numerosi casi ben documentati di studi riportati prematuramente, prima che fossero disponibili i dati completi, le cui conclusioni sono state successivamente smentite. In alcuni casi, gli studi sono stati presentati a congressi ma non successivamente pubblicati, o sono stati pubblicati con enorme ritardo. Ciò non è accettabile, perché non permette una valutazione indipendente tempestiva. Bisogna anche riconoscere che tutti gli studi che vengono fermati e pubblicati rapidamente per via di una grande differenza tra

i trattamenti, tendono ad esagerare il vero effetto. Questo è particolarmente vero quando gli sperimentatori sono tempestivi nel pubblicare lo studio in un momento in cui i risultati mostrano casualmente un'alta differenza, quando invece, se venissero accumulati risultati più sostanziali su più pazienti e follow-up, si potrebbe prevedere una «regressione verso la verità».

f. *Sono state prese misure per garantire che lo sponsor non fosse in grado di accedere ai dati non in cieco nel corso dello studio?* Il ruolo dello sponsor nello studio deve essere chiaro. C'è oggi un accordo generale sul fatto che lo sponsor (sia industriale che governativo) non dovrebbe essere al corrente dell'andamento dei risultati principali durante lo studio, dato il suo forte interesse personale. È necessario quindi fornire informazioni riguardo a chi ha gestito i dati, chi ha tenuto ai codici, chi aveva accesso ai dati non in cieco, e quando.

g. *È stato nominato un comitato indipendente di monitoraggio dei dati e della sicurezza?* È oggi pratica corrente che ci sia un comitato di monitoraggio dei dati e della sicurezza, formato di esperti delle situazioni cliniche in studio e da esperti in sperimentazione clinica e statistici. Questi esperti dovrebbero essere indipendenti dagli sperimentatori e dallo sponsor ed essere gli unici ad avere accesso ai dati nel corso dello studio. Dovrebbero condurre le analisi ad interim ed esprimere il proprio parere riguardo alla eventuale continuazione o interruzione dello studio per motivi di efficacia o sicurezza. Il rapporto dovrebbe descrivere la funzione e la composizione del comitato.

Risultati e conclusioni

a. *È stata impiegata una tecnica di analisi appropriata?* È ovviamente cruciale che l'analisi dei risultati dello studio venga condotta correttamente, utilizzando metodi statistici appropriati. Il rapporto dello studio dovrebbe presentare i risultati e definire i metodi statistici in modo da rassicurare il lettore riguardo alla correttezza dell'analisi e dell'interpretazione dei dati.

b. *Sono state riferite misure di variabilità, compresi i limiti fiduciarci?* L'obiettivo statistico di uno studio è in generale di fornire stime della grandezza della differenza tra i trattamenti per l'evento/i di interesse principale (per esempio il rischio relativo di mortalità o di un evento clinico, la differenza media di una variabile quantitativa). Il grado di incertezza di queste stime dipende dalla dimensione dello studio: più sono i

pazienti valutati, più è precisa la stima. È buona pratica statistica esprimere l'estensione dell'incertezza nei risultati, di solito includendo i limiti fiduciarci di ogni stima puntuale.

c. *Sono stati riportati i dati complessivi non aggiustati sulla misura dell'obiettivo principale e gli effetti collaterali importanti?* Mentre tecniche statistiche (talvolta complesse) sono utili per esprimere la dimensione e la significatività statistica delle differenze tra i trattamenti, è importante che al lettore del rapporto su uno studio clinico vengano forniti anche i dati di base. Quindi dovrebbero essere tabulati i numeri dei pazienti che in ciascun gruppo sono deceduti, hanno avuto eventi clinici ed effetti avversi, e dovrebbero essere fornite informazioni appropriate sulla variabilità (media e deviazione standard, o tabelle o grafici delle distribuzioni per gruppo di trattamento).

d. *È stato dichiarato il valore preciso della probabilità, e se il test di significatività era a una o a due code?* I rapporti sugli studi clinici spesso ipersemplicano i risultati dei test statistici di significatività, dichiarando $P < 0.05$ (significativo) o NS (non significativo), mentre è più corretto dal punto di vista statistico dichiarare il valore preciso di P . Infatti, la forza dell'evidenza di una differenza fra trattamenti è contenuta nel valore di P : ad esempio $P = 0.001$ è molto più convincente di $P = 0.04$, e $P = 0.06$ non è del tutto privo di interesse riguardo alla possibile esistenza di una differenza, mentre $P = 0.5$ significa che non ci sono indicazioni di una differenza vera.

e. *È stata data un'interpretazione appropriata delle differenze statisticamente non significative?* Un'interpretazione più che dogmatica dei test di significatività è un trabocchetto frequente in tutte le pubblicazioni mediche. In particolare, $P < 0.05$ non «prova» che i trattamenti sono differenti, e $P > 0.05$ non stabilisce che i trattamenti sono equivalenti. Più correttamente, più è piccolo il valore di P , più è forte l'evidenza che una differenza vera può esistere, e si dovrebbero evitare conclusioni eccessivamente dogmatiche basate sul raggiungimento (o meno) di un particolare livello di significatività statistica.

f. *È stata data un'enfasi corretta all'esposizione e all'interpretazione dell'analisi statistica, con particolare riguardo alle analisi non pianificate (ossia sottogruppi, obiettivi multipli, analisi multiple)?* La presentazione dei risultati dovrebbe essere coerente con gli obiettivi principali predefiniti dello studio e si dovrebbe evitare di enfatizzare risultati secondari «post hoc», specialmente se indicano un'ampia differenza tra i trattamenti. Per esempio, se un risultato principale dimostra una non evidenza di differenza, gli autori potrebbero essere tentati di cercare qualche

analisi secondaria che sia statisticamente significativa, per esempio in un particolare sottogruppo di pazienti o una misura differente della risposta dei pazienti o un periodo di tempo specifico per il follow-up, e costruire su questi punti conclusioni positive ingiustificabili.

g. *Le conclusioni riguardanti i confronti tra i trattamenti erano totalmente compatibili con gli obiettivi principali e i risultati presentati?* Yusuf e collaboratori,¹ nella discussione sull'analisi e l'interpretazione degli effetti del trattamento nei sottogruppi, fa riferimento a sottogruppi *propri* e *impropri*. Un *sottogruppo proprio* è caratterizzato da un insieme comune di parametri di base, come l'età, il sesso, una malattia specifica (per esempio un infarto pregresso, una sede particolare dell'infarto, una caratteristica elettrocardiografica specifica). Questi autori definiscono un *sottogruppo improprio* come un gruppo di pazienti caratterizzato da misure variabili dopo la randomizzazione e potenzialmente influenzabili dal trattamento, per esempio l'esame separato di pazienti con coronarie aperte in uno studio su un farmaco trombolitico, determinato dopo l'inizio del trattamento randomizzato. Aggiungono: «i confronti tra pazienti che rispondono alla terapia e quelli che non rispondono, oppure tra pazienti aderenti al trattamento e pazienti non aderenti, basati su informazioni raccolte dopo la randomizzazione, sono forme particolarmente significative di analisi di sottogruppi impropri, dato che la risposta e la aderenza possono essere marcatori di una buona prognosi, e non necessariamente misure di efficacia terapeutica».

Per motivi statistici, i sottogruppi da analizzare devono essere pochi, anche se sono stati identificati a priori. Come indicato in precedenza, qualsiasi analisi di sottogruppi non pianificata è fortemente sospetta, ma non è irragionevole analizzare i sottogruppi per cui c'è una plausibilità biologica, in particolare se la loro rilevanza è stata suggerita da altri studi. Per principio, l'analisi di sottogruppi dovrebbe essere considerata come generatrice di ipotesi da confermare in studi successivi, anche se tali studi sono raramente realizzabili. Naturalmente, il fatto che il problema dei sottogruppi sia discutibile sul piano statistico, non implica che non esistano sottogruppi con un comportamento differente riguardo all'effetto del trattamento.

Un altro problema è la presentazione dei risultati in termini di beneficio assoluto o di beneficio relativo acquisito. È molto comune purtroppo che vengano affermate riduzioni percentuali sicure di un evento-obiettivo. Questo è di scarso interesse clinico se il rischio è basso, anche se una riduzione relativa piccola può essere di grande interesse se il rischio è alto.

h. *Se il protocollo permetteva di somministrare il trattamento attivo in studio al gruppo di controllo in presenza di un'indicazione clinica, ne è stato tenuto conto nelle conclusioni?* Molti studi confrontano l'uso del trattamento da saggiare in tutti i pazienti reclutati nello studio, con l'uso selettivo del trattamento in presenza di un'indicazione clinica. Se ciò si traduce nella somministrazione del farmaco in studio a molti pazienti nel gruppo di controllo, in funzione delle convinzioni e della pratica clinica dei partecipanti, allora anche in uno studio molto grande può succedere di non riuscire a trarre conclusioni affidabili riguardo alla terapia in studio.

Dimensioni dello studio

Numerosità dei pazienti

Prescindendo dalla (non) significatività statistica, è evidente che lo studio è stato condotto su un numero troppo piccolo di pazienti per raggiungere conclusioni affidabili e generalizzabili? In ogni settore specifico della ricerca clinica cardiovascolare sono stati in generale acquisiti esperienza e buon senso, sufficienti a produrre un accordo generale su cosa sia un numero di pazienti e un follow-up adeguati in uno studio clinico. Purtroppo la maggior parte degli studi sono troppo piccoli e i lettori dovrebbero essere scettici riguardo alle loro conclusioni sia che esse siano positive, negative od equivoche.

Significatività statistica

Se lo studio è di piccole dimensioni, e, ciononostante, ha ottenuto risultati statisticamente significativi, sono stati presi in considerazione il «publication bias», i problemi di validità interna, e la possibilità che una differenza tra i trattamenti sia stata dilatata artificialmente? Perché uno studio piccolo possa produrre una differenza tra i trattamenti statisticamente significativa, la grandezza dell'effetto osservato deve essere grande, spesso in modo implausibile. Quindi, gli studi «statisticamente significativi» di piccole dimensioni tendono ad amplificare l'effetto reale (che potrebbe anche non esistere) e vanno interpretati con grande cautela. Inoltre, gli studi piccoli sono soggetti a «publication bias»: gli autori infatti non pubblicano gli studi piccoli «negativi», e

quindi la letteratura medica sugli studi piccoli è fortemente distorta verso un'amplificazione degli effetti veri e un alto rischio di risultati falso-positivi (errore di tipo I).

Mancanza di significatività statistica

Se uno studio non ha raggiunto una differenza convenzionalmente significativa, potrebbe essere successo perché è stato studiato un numero insufficiente di pazienti? Gli studi di piccole dimensioni non hanno una potenza statistica sufficiente a scoprire realisticamente differenze tra i trattamenti modeste ma importanti, e hanno quindi un rischio sostanziale di risultati falso-negativi (errore di tipo II). Quindi, è necessario cautelarsi nei confronti di conclusioni che affermano che non c'è un effetto quando lo studio aveva scarsa probabilità di cogliere questo effetto a causa di un numero insufficiente di pazienti. Ciò è particolarmente importante negli studi che si pongono come obiettivo una valutazione di equivalenza tra trattamenti, dove è troppo facile scivolare dal termine «evidenza insufficiente di una differenza» all'affermazione dogmatica illegittima «non esiste differenza».

Uso dei limiti fiduciari

Nel riassunto e nelle conclusioni dello studio è stata data enfasi sufficiente all'imprecisione delle stime? In un piccolo studio i limiti fiduciari di ogni differenza tra i trattamenti saranno molto ampi, e spesso includeranno effetti importanti ma implausibili ed effetti modesti (e, naturalmente un effetto nullo se la differenza non è statisticamente significativa). Quindi i limiti fiduciari rappresentano un valido strumento statistico per disarmare interpretazioni troppo ottimistiche di dati troppo scarsi.

Validità esterna

Pazienti studiati

a. I pazienti inclusi nello studio rappresentavano in modo adeguato i pazienti che si incontrano nella pratica clinica di routine? I criteri di inclusione erano troppo stretti o troppo ampi? Ciascuno studio dovrebbe

be indicare chiaramente i criteri di inclusione e di esclusione. In alcuni studi, le definizioni sono deliberatamente imprecise, in modo da poter includere tutti i pazienti affetti dalla malattia in studio. Sebbene tale definizione sia molto efficiente nel reclutare grandi numeri di pazienti, essa determina necessariamente l'inclusione di pazienti con altre diagnosi. Quando si impiegano criteri più restrittivi, una grande proporzione di pazienti viene esclusa. Un altro problema è relativo alla volontà degli sperimentatori di reclutare pazienti di vario tipo. Molti studi hanno ad esempio escluso deliberatamente le donne, ma potrebbero non averlo affermato con chiarezza. Nello stesso modo, c'è indubbiamente un «*bias*» di reclutamento degli anziani, anche se magari non sono esclusi per definizione.

Preoccupazioni riguardo ai rischi della terapia in studio possono tradursi in una tendenza a reclutare pazienti a relativamente basso rischio. Numerosi studi sulla terapia trombolitica hanno confermato questa impressione.

b. *Sono state prese misure adeguate ad assicurare che un'alta proporzione di pazienti eleggibili sia stata randomizzata? In particolare è stato tenuto un registro di tutti i pazienti con la stessa malattia e come è il confronto tra i pazienti randomizzati e quelli non randomizzati (sia eleggibili che non eleggibili)?* È decisamente meritevole tenere un registro dei pazienti con la malattia in studio, sia che abbiano o che non abbiano i criteri di inclusione, in modo da poter definire la proporzione e le caratteristiche dei pazienti reclutati. Quando questo è stato fatto, è stato possibile non solo determinare la percentuale della popolazione reclutata, ma anche scoprire in che misura i pazienti esclusi erano simili a quelli inclusi. In molti studi i pazienti esclusi hanno avuto un destino molto peggiore rispetto a quello dei pazienti assegnati al placebo o al gruppo di controllo dello studio. In generale, la mortalità del gruppo assegnato al placebo negli studi è più bassa di quanto previsto. Tuttavia i registri hanno un limite, determinato dalla difficoltà di definire chi debba essere in essi inserito. Le caratteristiche dei pazienti inclusi o esclusi nello studio possono variare fortemente da paese a paese e da centro a centro.

c. *Erano appropriati l'ambito dello studio e il modo di selezionare i pazienti? I criteri di eleggibilità erano troppo stretti o troppo ampi? Gli autori hanno estrapolato in modo inappropriato i loro risultati a tipi di pazienti non adeguatamente rappresentati nello studio? È frequente che gli autori generalizzino i loro risultati dai tipi limitati di pazienti inclusi nello studio a pazienti con la stessa patologia. Per esempio, è necessa-*

rio riconoscere che studi condotti in ambito ospedaliero non sono immediatamente estrapolabili alla medicina generale (i cui pazienti potrebbero essere meno gravi).

Trattamenti

a. *I trattamenti a confronto, compreso lo schema di dosaggio, la durata del trattamento, la non aderenza, il regime del gruppo di controllo (placebo o terapia standard) erano appropriati per la pratica clinica di routine e per determinare la futura strategia di trattamento in tali pazienti?* Nel valutare la rilevanza (o mancanza) di qualsiasi differenza tra i trattamenti è necessario osservare le strategie attive e di controllo come realmente implementate nello studio e considerare la loro applicabilità alla pratica clinica di routine. Se il controllo era un placebo, ma altri trattamenti attivi hanno dimostrato un'evidenza di effetto in passato, allora lo studio non può giustificare direttamente l'uso del trattamento rispetto a queste alternative. Se la durata del trattamento è stata relativamente breve allora l'estrapolazione a periodi di trattamento più lunghi in una condizione cronica può non essere giustificata. Se mancano dettagli sull'uso reale del farmaco (ossia deviazioni dalla dose prevista, specialmente se dovute ad eventi avversi) allora l'applicabilità ai pazienti futuri può essere poco chiara.

b. *Sono stati presi in considerazione in modo adeguato tutti gli aspetti della buona pratica clinica corrente (per esempio la cura di supporto)?* È importante che l'uso di altri trattamenti (farmaci concomitanti alla terapia randomizzata) sia specificato chiaramente, sia in termini di criteri di inclusione/esclusione che di numeri concreti di pazienti che hanno ricevuto le terapie concomitanti. Lo scopo è di rendere chiaro il ruolo di un nuovo trattamento nel contesto degli altri trattamenti esistenti. In particolare se un nuovo farmaco è visto come un potenziale supplemento ad altre terapie stabilite, è necessario valutare il possibile rischio di interazioni avverse da terapie multiple.

Misure di «outcome» e follow-up

a. *Le misure di «outcome» (obiettivi dello studio, indicatori di risposta del paziente) erano appropriate per raggiungere le conclusioni riguardo ai trattamenti in studio?* In condizioni in cui è elevata, la mortalità può essere l'unico obiettivo principale appropriato, ma quando la

mortalità da sola non è un obiettivo soddisfacente, può essere combinata con altri eventi. Questi «endpoint combinati» possono porre problemi. Per esempio, la mortalità può risultare aumentata e l'infarto miocardico diminuito. In alcuni studi sull'insufficienza cardiaca la qualità della vita è risultata migliorata, ma la mortalità è risultata aumentata. L'interpretazione di questi risultati è chiaramente questione di opinioni più che di scienza.

b. *È stata data troppa importanza agli indicatori surrogati di risposta (per esempio indicatori fisiologici) piuttosto che agli indicatori principali di prognosi complessiva (per esempio mortalità, eventi clinici maggiori)?* Come conseguenza del fatto che gli «hard endpoints» come la morte e l'infarto miocardico possono essere rari, si è sviluppata l'infelice tendenza ad affidarsi ad endpoint surrogati come la pervietà delle coronarie per gli studi sui farmaci fibrinolitici, la frazione di eiezione per gli studi sullo scompenso cardiaco, e la soppressione delle aritmie per gli studi sui farmaci antiaritmici. Come Lipicky e Packer² hanno sottolineato, per utilizzare una variabile fisiologica come surrogato ci dovrebbe essere non solo una relazione statistica tra l'endpoint proposto e l'outcome clinico, ma dovrebbe anche esistere una base patofisiologica per essere certi che il surrogato sia il determinante principale dell'outcome. Tuttavia, anche questo non è abbastanza, perché anche se la variabile è il determinante principale in circostanze normali, ciò può non essere acquisito nel contesto sperimentale. Inoltre, il tentativo di limitare la dimensione dello studio con l'uso delle variabili surrogate, può nascondere alcuni effetti avversi che possono essere visti solo con uno studio di grandi dimensioni.

c. *La durata del trattamento e la lunghezza del follow-up erano sufficientemente affidabili per valutare l'efficacia e la sicurezza della terapia?* La durata del follow-up negli studi clinici è molto variabile. Ciò può avere conseguenze pesanti sulla valutazione del trattamento. Alcuni studi nell'infarto miocardico si sono fermati alla mortalità della prima settimana o fino alla dimissione dall'ospedale. Più comunemente, il follow-up è di 4-6 settimane, ma i medici hanno ovviamente la necessità di conoscere l'effetto sulla prognosi su un periodo di tempo più lungo, ad esempio almeno un anno. In alcuni casi un follow-up più lungo dimostra che il trattamento è più efficace di quanto apparso in un primo tempo. Nel caso della terapia ipolipemizzante, non ci si attende alcun beneficio per almeno due anni, e si potrebbe concludere che tutti gli studi dovrebbero avere un follow-up di almeno 5 anni. In termini di sicurezza è desiderabile un follow-up più lungo, dato che possono pas-

sare molti anni prima che gli effetti avversi si manifestino, come si è visto con la terapia ormonale sostitutiva.

d. *Sono state prese misure adeguate per far emergere tutti gli eventi avversi e gli effetti collaterali rilevanti del trattamento?* Dovrebbe essere fornita una tabella con gli effetti avversi incontrati, e la relazione dovrebbe spiegare come sono stati raccolti i dati relativi agli eventi avversi.

Discussione equilibrata e conclusioni

a. *Sono stati considerati adeguatamente da parte degli autori i limiti del loro studio rispetto a quanto fino qui riportato?* Tutti gli articoli dovrebbero includere una discussione sui limiti dello studio che viene presentato. Questa sezione può rivelare aspetti dello studio che possono non essere stati evidenti per il lettore e dà modo di comprendere l'oggettività degli autori.

1. *Titolo dell'articolo.* Il titolo è importante, perché spesso è la prima cosa che attrae l'attenzione. Non dovrebbe fare affermazioni eccessive e dovrebbe riflettere accuratamente il contenuto.
2. *Riassunto.* Molti lettori non vanno al di là del riassunto. È quindi essenziale che il riassunto rifletta il contenuto dell'articolo, e che faccia affermazioni caute.
3. *Materiali e metodi dello studio.* L'articolo dovrebbe descrivere brevemente l'ipotesi che è stata valutata, i principi statistici con cui è stato disegnato, con il calcolo della potenza, i criteri di esclusione e di inclusione, la popolazione studiata con le definizioni dei problemi clinici e degli obiettivi studiati, il metodo di randomizzazione, la gestione dei dati, il modo in cui sono state mantenute le condizioni di cecità, il ruolo dei comitati, il ruolo degli sponsor.
4. *Discussione.* Nella discussione gli autori in genere rivedono i loro risultati e le implicazioni del loro studio per il tipo di pazienti inclusi. Ma per suscitare un più ampio interesse dovrebbero considerare quanto le loro conclusioni possono essere trasferite a una popolazione più ampia di pazienti simili. Devono tenere presente che, per quanto universali fossero i criteri di inclusione, i pazienti reclutati nello studio sono una popolazione selezionata. Gli studi clinici sono condotti in circostanze piuttosto artificiali, con un'attenzione maggiore dell'usuale, con medici e infermieri dedicati, con pazienti ade-

renti al trattamento. Spesso hanno incluso, o deliberatamente o a causa delle circostanze, un gruppo veramente ristretto di pazienti, come per esempio, solo, o in modo predominante, bianchi, solo uomini, solo scandinavi. Estrapolazioni valide ad altri gruppi dipendono dalla forza statistica dell'evidenza, dalla plausibilità biologica dell'estrapolazione, e dall'evidenza disponibile da altre fonti (sia sperimentali che osservazionali).

b. *Gli autori hanno descritto in modo equilibrato le evidenze provenienti da altri studi o hanno dato un peso sproporzionato ai loro risultati?* È essenziale che gli autori considerino i loro risultati alla luce di altre pubblicazioni e riconoscano il dovuto merito agli altri ricercatori del settore. Se l'effetto del trattamento stimato da loro è (o non è) compatibile con studi precedenti correlati, allora l'evidenza comparativa quantitativa dovrebbe essere documentata insieme con le interpretazioni e i motivi di incoerenza (comprese le differenze nel disegno dello studio e il gioco del caso).

Lo scopo dovrebbe essere quello di presentare con chiarezza qualsiasi nuovo importante risultato di uno studio, coerentemente con le ipotesi principali definite in precedenza, di discutere apertamente i limiti dello studio, di interpretare le nuove acquisizioni nel contesto dei dati rilevanti già esistenti, in modo che sia gli sperimentatori clinici che pianificano future ricerche che i medici responsabili della cura dei futuri pazienti possano prendere decisioni razionali basate su evidenze oggettive.

Bibliografia

1. Yusuf S, Wittes J, Probstfield J, Tyroler HA. Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. *JAMA* 1991; 266: 93-8.
2. Lipicky RL, Packer M. Role of surrogate end-points in the evaluation of drugs for heart failure. *J Am Coll Cardiol* 1993; 22 (Suppl A); 179A-84A.