

La valutazione della qualità degli studi clinici

Note per un medico che interpreta un trial clinico o una meta-analisi

Tre precedenti articoli di “ABC degli studi clinici” hanno avuto come oggetto le sperimentazioni controllate e randomizzate (trial, RCTs) e la meta-analisi (1-3). Il presente articolo riprende i contenuti di quegli articoli e si concentra sui criteri applicabili per giudicare la credibilità di un trial o di una meta-analisi, con particolare attenzione a quelli più utili dal punto di vista della pratica medica.

1. Trial controllati e randomizzati

- a. Una prima difficoltà consiste nel fatto che la credibilità di un trial dev'essere giudicata sul lavoro pubblicato, ed è possibile che un trial ben disegnato e ben condotto sia riportato in modo inadeguato, o viceversa, che deficienze metodologiche del trial non emergano dallo scritto. Nel 1996 è stato raccomandato uno schema di esposizione dei trial (CONSORT: *Consolidation of the Standards of Reporting Trials* (4,5)), con lo scopo di assicurare ai lettori la possibilità di trovare nella pubblicazione le informazioni necessarie a interpretare il trial e a valutarne la credibilità. Alcune riviste esigono per pubblicare un trial che l'esposizione segua le norme del CONSORT (ad esempio *JAMA*, *BMJ*, *Lancet*, *CMAJ*), e questo può essere una forma di difesa per il medico che interpreta e valuta un trial. E' chiaro, però, che questo garantisce – e fino a un certo punto - la qualità di esposizione del trial, e non la sua credibilità.
- b. Qualche volta vengono pubblicati (e ci si chiede come siano stati accettati) trial con un campione particolarmente sottodimensionato. Un esempio limite è un trial dell'interferone beta per la prevenzione delle recidive post-resezione del carcinoma primitivo del fegato, con 10 pazienti trattati e 10 controlli (6). E' intuitivo che la randomizzazione non può garantire che due gruppi così piccoli siano clinicamente comparabili, con la conseguenza che viene violato il principio base del trial randomizzato.
- c. Qualche volta un trial che si conclude con un risultato di beneficio terapeutico del trattamento sperimentale riporta un andamento dei controlli particolarmente sfavorevole, che non è in linea con altri studi precedenti e successivi, randomizzati o di coorte. Un esempio è il trial citato nel paragrafo precedente (6), nel quale, in due anni, sette su dieci controlli hanno presentato una recidiva del tumore con un'incidenza all'incirca doppia rispetto a quella abitualmente

osservata; un altro esempio è un trial di scleroterapia profilattica delle varici esofagee nei cirrotici (7), nel quale i controlli non sclerosati avevano una incidenza di emorragie doppia rispetto alla mediana osservata in altri 18 trial e in numerosi studi di coorte (8). Ne deriva che il medico ha bisogno di conoscere la storia naturale (o la storia post-terapia standard) di una condizione clinica per valutare la validità esterna di un trial. Sarebbe infatti poco saggio applicare ai propri pazienti un trattamento apparso vantaggioso in pazienti con aspettative prognostiche inusualmente sfavorevoli.

- d. Si trovano in letteratura numerosi esempi di trial randomizzati di uno stesso trattamento applicato alla stessa malattia. Spesso i risultati di questi trial sono omogeneamente positivi (o negativi), come ci si attende, e questa è una conferma incrociata dei rispettivi risultati. Non sono pochi, però, i trial con risultati discordanti (alcuni favorevoli al trattamento sperimentale, altri senza differenze significative fra trattamento sperimentale e controllo, altri favorevoli al controllo). Alcuni esempi: scleroterapia profilattica nella cirrosi con varici esofagee (8), antibiotici nella tosse acuta (9), calcio per la prevenzione della preeclampsia (10), trattamenti eradicanti dell'*Helicobacter pylori* nella dispepsia (11). Molte di queste discordanze si spiegano con differenze delle misure di efficacia dei trattamenti (9,12) o con differenti caratteristiche cliniche dei pazienti studiati, dimostrate (8,10) o ipotizzabili (11,12). Altre non sono facilmente spiegabili. L'esistenza di trial discordanti è comunque un'insidia per il medico che potrebbe avere conoscenza soltanto di uno di essi (positivo o negativo), ed è un elemento a favore delle meta-analisi come guida per la medicina pratica; la meta-analisi infatti rende note le discordanze, e induce (o dovrebbe indurre) ad un atteggiamento clinico dubitativo.
- e. E' necessario che i trial relativi a malattie a decorso cronico misurino come indici di efficacia dei trattamenti il loro effetto su end point surrogati. Valgano come esempio i trial dei trattamenti dell'epatite cronica C in soggetti non cirrotici, o dell'ipertensione arteriosa moderata: il tempo necessario per la comparsa di manifestazioni cliniche (scompenso della cirrosi nel primo caso; *stroke* nel secondo) o per la mortalità è infatti misurabile in decenni. E' però importante che gli end point surrogati scelti siano in qualche modo correlati con gli end point clinicamente rilevanti: per esempio, nelle epatiti croniche, è ben

dimostrato che l'eradicazione dell'infezione perdurante oltre 6 mesi dopo la fine del trattamento è seguita dalla stabile regressione dei segni umorali e istologici di malattia, cioè dalla guarigione (13).

I paragrafi precedenti introducono la presentazione e un breve commento dell'indice di qualità dei trial più diffusamente adottato, che è quello di Jadad (14), di seguito riportato:

- **Lo studio è descritto come randomizzato?**
- **Lo studio è descritto come in doppio cieco?**
- **C'è una descrizione delle sospensioni di trattamento e dei drop out?**

Si attribuisce **1** punto per la risposta positiva a ognuna delle tre domande; si attribuiscono **0** punti per la risposta negativa.

Si esamina il testo: si aggiunge **1** punto alla randomizzazione e **1** punto alla cecità se sono metodologicamente appropriate, e si sottrae **1** punto alla randomizzazione e **1** punto alla cecità se inappropriate.

Il "punteggio" massimo possibile è 5; il trial si giudica di cattiva (*poor*) qualità se il "punteggio" è inferiore a 3.

L'indice di Jadad richiede tuttavia qualche chiarimento.

"Appropriata" è la randomizzazione che renda impossibile per gli sperimentatori prevedere l'assegnazione dei pazienti al trattamento sperimentale o a quello di controllo. Il modo migliore di acquisire questa garanzia è di effettuare la randomizzazione a distanza, di solito per telefono, delegandola a un centro diverso da quelli che direttamente conducono il trial.

La cecità è assolutamente necessaria quando gli end point implicano un margine di soggettività – per esempio l'interpretazione di *imaging*, come ecografie, radiografie o TAC, oppure l'intensità o la presenza di sintomi come il dolore. Lo è meno per end point inequivocabili (la morte); qualunque sia l'end point, la cecità è necessaria nella valutazione finale dei risultati del trial.

La registrazione e la comunicazione delle sospensioni del trattamento e dei *drop out* (cioè dei pazienti di cui si sono perse le tracce) forniscono informazioni essenziali sulla tollerabilità del trattamento e sulla capacità degli sperimentatori di mantenere la *compliance* dei pazienti. Trial con oltre il 15-20% di sospensioni e *drop out* sono considerati con molta cautela.

Infine, è opportuno ricordare due "punti deboli" dei trial.

– I trial sono relativamente poco sensibili al rilievo delle reazioni avverse ai farmaci, la cui identificazione e stima di incidenza derivano più spesso da segnalazioni e studi post-commercializzazione (15,16). Alcuni esempi (mibefradil, alendronato) sono stati riportati in un precedente articolo di questo ABC (2); altri esempi riguardano i farmaci anoressizzanti fenfluramina e dexfenfluramina, ritirati dal commercio nel 1997 perché responsabili di valvulopatie e ipertensione polmonare (17,18), non rilevate in precedenti trial (19); e gli antiaggreganti tienopiridinici ticlopidina e clopidogrel, associati con il rischio di porpora trombotica trombocitopenica (20,21) e di sindrome emolitico-uremica (22), non rilevato nei trial pre-commercializzazione (23,24).

– Un problema difficilmente risolvibile riguarda le influenze promozionali dell'industria, che sponsorizza la stragrande maggioranza dei trial. Malgrado tali influenze non siano facilmente rilevabili, è stato comunque osservato che i trial sponsorizzati dalle industrie farmaceutiche evidenziano con singolare frequenza risultati favorevoli ai farmaci sperimentali prodotti da quelle industrie (25-28); inoltre sono stati dimostrati rapporti di collaborazione fra accademici che hanno pubblicato studi o commenti favorevoli a un farmaco o a una classe di farmaci e le ditte produttrici degli stessi (29, 30).

2. Meta-analisi

- a. Le meta-analisi presentano la sintesi quantitativa dei risultati di trial controllati e randomizzati dello stesso trattamento in una particolare malattia o condizione clinica. Molte meta-analisi scelgono di basarsi sui trial pubblicati in *extenso* (dei quali è possibile fare una valutazione qualitativa), altre includono anche i trial pubblicati come *abstract* (ottenendo una più completa visione dei risultati del trattamento), altre – più rare – tentano di localizzare trial non ancora pubblicati. Dal punto di vista del medico, la meta-analisi ha il vantaggio di partire da una ricerca il più possibile esauriente e idealmente completa di tutti i trial di un determinato trattamento. Il medico evita così il rischio di doversi basare su uno solo di più trial disponibili, ignorando l'eventuale sistematicità degli altri.
- b. Tuttavia, il medico di solito non è in grado di verificare se una meta-analisi include realmente tutti i trial disponibili; pertanto, non è in grado di applicare questo criterio di validità delle meta-analisi. Deve allora confidare che questa verifica sia stata realizzata dal Comitato Editoriale che ha approvato la pubblicazione della meta-analisi. Da questo punto di vista (e da altri, come si vedrà) sono particolarmente affidabili le meta-analisi della *Cochrane Collaboration* (v. in 3), disponibili via Internet o in abbonamento a CD-ROM periodici. Ogni meta-analisi della *Cochrane* è infatti preceduta dall'esame preliminare del protocollo ad opera di comitati di esperti per area di patologia (ad es. cardiologi, reumatologi, epatologi, ecc.), che eventualmente rigettano il protocollo o ne condizionano l'approvazione a modifiche anche radicali; lo stesso processo di valutazione, approvazione, approvazione con modifiche o rigetto si ripete per la bozza finale della meta-analisi. Questa metodologia consente di elaborare meta-analisi di qualità superiore alla media di quelle pubblicate (31). Le meta-analisi *Cochrane*, tuttavia, sono disponibili solo per una parte dei trattamenti di impiego corrente.

c. Le discordanze fra i risultati di più trial di uno stesso trattamento (v. sopra, paragrafo 1d) si trasferiscono nelle meta-analisi costruite su quei trial. Può accadere, cioè, che in una meta-analisi la direzione dell'effetto terapeutico sia favorevole al trattamento sperimentale in alcuni trial, e favorevole al trattamento di controllo in altri. Un esempio particolarmente evidente è riportato nella Figura 5 dell'articolo di "ABC degli studi clinici" riguardante la meta-analisi (3). Altri esempi si trovano nel paragrafo 1d di questo stesso articolo, e nella bibliografia in esso citata. Dal punto di vista pratico, il problema si può tradurre in una domanda: cosa fare quando una meta-

analisi si conclude con un risultato favorevole al trattamento sperimentale, ma include anche trial con risultati favorevoli al trattamento di controllo (Figura 1; v. anche voce bibliografica 8 e Figura 5 della voce bibliografica 3)? Un orientamento prudente suggerisce in questi casi di non applicare il trattamento sperimentale, anche se il risultato cumulativo è favorevole e statisticamente significativo. Non è possibile infatti prevedere se il paziente a cui si vuol applicare il trattamento se ne avvantaggerà, come quelli dei trial con risultati favorevoli, o ne sarà danneggiato, come quelli dei trial con risultati sfavorevoli.

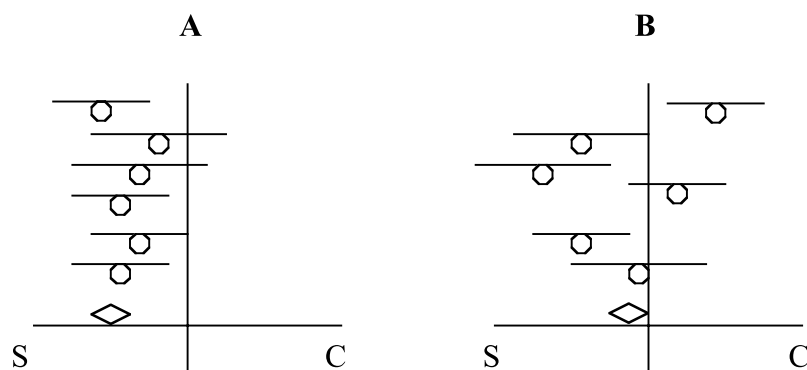


Figura 1. Due esempi di meta-analisi con risultato cumulativo (\diamond) favorevole al trattamento sperimentale e statisticamente significativo. **A:** direzione dell'effetto terapeutico omogeneamente favorevole al trattamento sperimentale (S) in tutti i trial; **B:** direzione dell'effetto terapeutico eterogenea, in alcuni trial favorevole al trattamento sperimentale (S), in altri al trattamento di controllo (C). Nel secondo esempio, un atteggiamento prudente suggerirebbe di non applicare il trattamento sperimentale, nonostante il risultato cumulativo favorevole e statisticamente significativo.

d. In una meta-analisi, il campione è ovviamente pari al numero totale dei pazienti dei trial in essa inclusi; si eleva di conseguenza la potenza delle meta-analisi, così che differenze relativamente modeste fra trattamenti sperimentali e controlli raggiungono la significatività statistica. Se a ciò si aggiunge il *publication bias* che riduce la probabilità di pubblicazione dei trial negativi, appare opportuno usare cautela nel trasferire alla pratica meta-analisi che si concludono con un risultato cumulativo solo marginalmente favorevole. Un esempio di meta-analisi con risultato statisticamente significativo a favore del trattamento sperimentale, ma di significatività clinica assai dubbia, è quello relativo alla terapia eradicante dell'*Helicobacter pylori* nella dispepsia non ulcerosa (32). Si tratta di una meta-analisi che include 9 trial del trattamento eradicante contro placebo, con risultato cumulativo a un anno significativamente favorevole al trattamento sperimentale. Un'analisi più ravvicinata rileva però che, posto a 100 il rischio dei controlli ancora dispeptici dopo un anno, il rischio relativo di dispepsia non modificata dei pazienti eradicati era di 91 (dispepsia regredita o migliorata nel 28% dei controlli e nel 36% degli eradicati). La significatività del risultato ($p < 0.0002$) deriva essenzialmente dalla grandezza del campione (2.541 pazienti).

Bibliografia

1. ABC degli studi clinici. 3. Sperimentazioni controllate e randomizzate. Aspetti metodologici. *BIF* 1999;4:40-2.
2. ABC degli studi clinici. 3. Sperimentazione controllata e randomizzata. Aspetti etici e note per la interpretazione degli RCTs. *BIF* 2000;1:33-8.
3. ABC degli studi clinici. 4. Meta-analisi. *BIF* 2000;2:36-42.
4. Begg C et al. Improving the quality of reporting of randomized controlled trials: the CONSORT statement. *JAMA* 1996;276:637-9.
5. Altman DG. Better reporting of randomized controlled trials: the CONSORT statement. *BMJ* 1996;313:570-1.
6. Ikeda K et al. Interferon beta prevents recurrence of hepatocellular carcinoma after complete resection or ablation of the primary tumor – a prospective randomized study of hepatitis C virus-related liver cancer. *Hepatology* 2000;32:228-32.
7. Paquet HJ. Prophylactic endoscopic sclerosing treatment of the esophageal wall in varices: a prospective controlled randomized trial. *Endoscopy* 1982;14:4-5.
8. Pagliaro L et al. Prevention of first bleeding in cirrhosis. A meta-analysis of randomized trials of nonsurgical treatment. *Ann Intern Med* 1992;117:59-70.
9. Fahey T et al. Quantitative systematic review of randomized controlled trials comparing antibiotic with placebo for acute cough in adults. *BMJ* 1998;316:906-10.
10. DerSimonian R, Levine RJ. Resolving discrepancies between a meta-analysis and a subsequent large controlled trial. *JAMA* 1999;282:664-70.
11. Danesh J, Pounder RE. Eradication of *Helicobacter pylori* and non ulcer dyspepsia. *Lancet* 2000;355:766-7.

12. Lindbaek M, Hjortdahl P. How do two meta-analyses of similar data reach opposite conclusions? *BMJ* 1999;**318**:873-4.
13. EASL International Consensus Conference on Hepatitis C. Paris, 26-28 February 1999. Consensus Statement. *J Hepatol* 1999;**30**:956-61.
14. Jadad A. Assessing the quality of RCTs: why, what, how, and by whom? In: Randomised Controlled Trials. London, BMJ Books 1998:45-60.
15. Wood AJJ. Thrombotic thrombocytopenic purpura and clopidogrel – a need for new approaches to drug safety. *N Engl J Med* 2000;**342**:1824-6.
16. Kaufman DW, Shapiro S. Epidemiological assessment of drug-induced disease. *Lancet* 2000;**356**:1339-43.
17. Connolly HM. Valvular Heart Disease associated with fenfluramine-Phentermine. *N Engl J Med* 1997;**337**:581-8.
18. Mark EJ et al. Fatal pulmonary hypertension associated with short-term use of fenfluramine-Phentermine. *N Engl J Med* 1997;**337**:602-6.
19. Guy-Grand B et al. International trial of long term Dexfenfluramine in obesity. *Lancet* 1989;**2**:1142-4.
20. Bennett CL et al. Thrombotic thrombocytopenic purpura associated with ticlopidine: a review of 60 cases. *Ann Intern Med* 1998;**128**:541-4.
21. Bennett CL et al. Thrombotic thrombocytopenic purpura associated with clopidogrel. *N Engl J Med* 2000;**342**:1773-7.
22. Oomen PHN et al. Hemolytic uremic syndrome in a patient treated with clopidogrel. [Letter] *Ann Intern Med* 2000;**132**:1006.
23. Hass WK et al. A randomized trial comparing ticlopidine hydrochloride with aspirin for the prevention of stroke in high risk patients. *N Engl J Med* 1989;**321**:501-7.
24. Caprie Stering Committee. A randomised blinded trial of clopidogrel vs aspirin in patients at risk of ischaemic events (CAPRIE). *Lancet* 1996;**348**:1333-8.
25. Bero LA et al. The publication of sponsored symposiums in medical Journals. *N Engl J Med* 1992;**327**:1135-40.
26. Cho MK, Bero LA. The quality of drug studies published in symposium proceedings. *Ann Intern Med* 1996;**124**:485-9.
27. Djulbegovic B et al. The uncertainty principle and industry-sponsored research. *Lancet* 2000;**356**:635-8.
28. Bero LA, Rennie D. Influences on the quality of published drug studies. In *J Technol Assess Health Care* 1996;**12**:209-37.
29. Stelfox HT et al. Conflict of interest in the debate over calcium channel antagonists. *N Engl J Med* 1998;**338**:101-6.
30. Bodenheimer T. Uneasy alliance. Clinical investigators and the pharmaceutical industry. *N Engl J Med* 2000;**342**:1539-44.
31. Jadad AR et al. Methodology and reports of systematic reviews and meta-analyses. A comparison of Cochrane reviews with articles published in paper-based journals. *JAMA* 1998;**280**:278-80.
32. Moayyedi P et al. Systematic review and economic evaluation of Helicobacter pylori eradication treatment for non-ulcer dyspepsia. *BMJ* 2000;**321**:659-64.

Prescrivere in base ai numeri

Interpretare con obiettività i dati numerici degli studi clinici – e i risultati che ne derivano – richiede una valutazione attenta del loro significato in relazione a diversi fattori, quali il numero di soggetti coinvolti, l'importanza e la gravità delle patologie considerate e degli eventi misurati nonché la rilevanza in termini di trasferibilità degli effetti osservati.

Non sempre, tuttavia, i risultati vengono espressi utilizzando gli indicatori che possono descrivere in modo più immediato l'effettiva validità del trattamento sperimentato. Per evitare di enfatizzare certi risultati è opportuno saper derivare dai dati presentati gli indicatori più idonei per una valutazione più completa e conseguentemente per un'adeguata prescrizione.

I risultati degli studi clinici sono spesso presentati in termini di **riduzione relativa di rischio (RRR)** di eventi sfavorevoli osservata in un gruppo di pazienti sottoposti a trattamento in sperimentazione (**EER** = Experimental Event Rate) rispetto a un gruppo di controllo (**CER** = Control Event Rate). La riduzione del rischio relativo, di solito espressa in valore percentuale, è un indicatore che, se non completato da ulteriori elementi di valutazione, può enfatizzare l'efficacia del trattamento al di là del suo significato clinico reale. L'esempio di seguito riportato ne dà spiegazione.

Supponiamo che, in uno studio, 10.000 pazienti siano trattati con un farmaco sperimentale e altrettanti (gruppo di controllo) con placebo, e che siano complessivamente evidenziati 1.000 eventi sfavorevoli nel gruppo sperimentale e 2.000 nel gruppo di controllo. La EER è pari al 10% e la CER al 20%. La RRR nel gruppo trattato è del 50%.

In un'altra ricerca clinica, in cui uno stesso numero di pazienti è stato sottoposto a trattamento sperimentale o di controllo, si osservano rispettivamente 100 (EER = 1%) e 200 (CER = 2%) eventi sfavorevoli. Anche in

questo caso la RRR nel gruppo trattato è del 50%.

Lo stesso valore RRR (50%) si ottiene se la EER è pari allo 0,1% (10 casi) e la CER allo 0,2% (20 casi) oppure se è la EER è 0,01% (1 evento) e la CER 0,02% (2 eventi).

Da tutto ciò deriva che, se la RRR è enucleata dall'incidenza reale dell'evento che in una data condizione clinica si desidera prevenire, è limitatamente significativa e poco utile al medico, in quanto lo può indurre a una sovrastima dell'impatto terapeutico.

La **riduzione assoluta del rischio (ARR)** è la differenza tra la quota di eventi osservati nel gruppo sperimentale e nel gruppo di controllo. Nei quattro casi citati, mentre la RRR è sempre del 50%, la ARR è, in termini percentuali, rispettivamente del 10%, 1%, 0,1%, 0,01%. Detto in altro modo, rispetto al gruppo di controllo, nel primo esempio su 100 pazienti trattati 10 presentano una riduzione del rischio di eventi sfavorevoli; negli altri tre casi, 1 paziente vedrà il rischio di eventi ridotto rispettivamente ogni 100, 1.000, 10.000 pazienti trattati.