Come viene riportata l'importanza clinica dei risultati degli studi?

Analisi di alcune sperimentazioni controllate e randomizzate

How well is the clinical importance of study results reported? An assessment of randomized controlled trials. Chan KBY et al. *CMAJ 2001;**165**:1197-1202.*

PARTE PRIMA

Riassunto

Contesto

L'interpretazione dei risultati delle sperimentazioni controllate e randomizzate (RCTs, dall'inglese *randomized controlled trials*) è basata sulla significatività statistica piuttosto che sull'importanza clinica. Il nostro obiettivo è stato valutare la qualità del modo in cui vengono riportati i fattori connessi all'importanza clinica in un campione di RCT pubblicati.

Metodi

In un campione random di 27 (su un totale di 266) RCT pubblicati in cinque delle maggiori pubblicazioni medico-scientifiche nel corso di un anno, quattro revisori indipendenti hanno analizzato i fattori considerati importanti per l'interpretazione dell'importanza clinica dei risultati degli studi: identificazione di un esito primario chiaramente definito; modalità con cui è riportata la differenza attesa tra gruppi usata per il calcolo della numerosità del campione (valore delta), e se tale differenza era basata sulla minima differenza di intervento clinicamente importante; significatività statistica dei risultati, presentazione di intervalli di confidenza pertinenti; interpretazione dell'importanza clinica dei risultati secondo gli autori.

Risultati

Ventidue dei 27 (81%) articoli riportavano esplicitamente un singolo esito primario. Dei 20 articoli che prevedevano un calcolo della numerosità del campione, 18 (90%) riportavano un valore delta. Due dei 18 (11%) articoli affermavano esplicitamente che il valore delta era scelto in modo da riflettere la minima differenza di intervento clinicamente importante. Per quanto riguarda gli esiti primari, gli intervalli di confidenza per le stime dell'efficacia degli interventi erano riportati in 11 dei 27 (41%) studi. I risultati dello studio erano interpretati dal punto di vista dell'importanza clinica in 20 articoli su 27 (74%). Di questi 20 articoli, 5 (25%) fornivano una giustificazione dell'interpretazione clinica dei risultati.

Interpretazione

Gli Autori di RCT pubblicati sulle maggiori riviste di medicina generale e di medicina interna non sempre forniscono la loro interpretazione dell'importanza clinica dei risultati, e spesso non danno informazioni sufficienti a consentire ai lettori di elaborare una propria interpretazione.

L'interpretazione dei risultati degli RCT è concentrata sulla loro significatività statistica piuttosto che sull'importanza clinica. Per esempio, la recente revisione dei CONSORT ("*Consolidated Standards Of Reporting Trials*": una serie di raccomandazioni largamente seguite e finalizzate a migliorare la qualità delle pubblicazioni degli RCT) non raccomanda specificamente che gli autori discutano l'importanza clinica dei loro risultati (1). La mancanza di attenzione all'importanza clinica dei risultati degli RCT ha condotto a errori e discordanze nella loro interpretazione, e alla tendenza a considerare equivalenti significatività statistica e importanza clinica. In un certo numero di circostanze, risultati statisticamente significativi possono essere clinicamente non importanti e, viceversa, risultati statisticamente non significativi non escludono la possibilità di effetti clinicamente importanti (2,3).

La differenza minima clinicamente importante (MCID, dall'inglese *minimal clinically important difference*) fra una terapia sperimentale e la rispettiva terapia di controllo è definita come quella differenza che sarebbe sufficiente a giustificare un cambiamento nelle decisioni terapeutiche, tenendo anche conto del rischio di eventi avversi, inconvenienti e costi (4). La MCID è un concetto chiave sia nel disegno che nell'interpretazione dei risultati degli RCT. Nel disegno, infatti, la numerosità del campione (*sample size*) del trial dovrebbe riflettere la MCID che si vuole evidenziare fra il trattamento in sperimentazione e quello di controllo (valore delta). Nell'interpretazione, il raggiungimento o meno di una MCID è il criterio principale di cui tener conto per la prospettiva di applicare i risultati del trial alle decisioni terapeutiche.

Ad esempio, in soggetti senza precedenti di infarto miocardico (MI, dall'inglese *Myocardial Infarction*) o ictus, il regolare impiego di aspirina riduce l'incidenza di MI da 0,70 a 0,50 per cento e per anno (riduzione di rischio: assoluta di 0,20%, relativa di circa il 25%), ma questo vantaggio potrebbe essere controbilanciato da un concomitante aumento di incidenza di ictus da 0,30 a 0,32 per cento e per anno (aumento di rischio: assoluto di 0,02%, relativo di circa il 10%), e di emorragia gastrointestinale da 1 a 2 per cento e per anno (5). Dopo avere valutato vantaggi e svantaggi dell'uso dell'aspirina in questo contesto, un autorevole *panel* di esperti non ne ha raccomandato l'uso, giudicando che la sua efficacia nel prevenire l'MI non era sufficiente a superare l'aumento

di incidenza di ictus e di emorragia gastrointestinale. Cioè, in questo contesto l'efficacia dell'aspirina non era sufficiente a raggiungere o a superare la MCID.

PARTE SECONDA

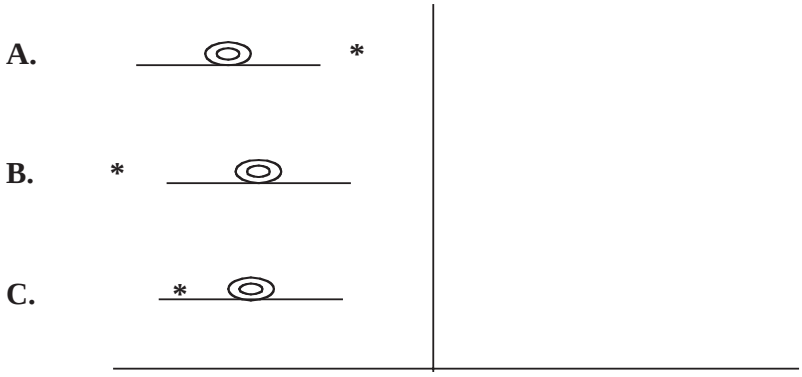
Lo studio ha randomizzato il 10% dei 266 RCT pubblicati tra il 1 dicembre 1998 e il 1 novembre 1999 da cinque riviste di medicina generale con il massimo *impact factor* (*Annals of Internal Medicine, British Medical Journal, Journal of the American Medical Association, The Lancet, New England Journal of Medicine*). La traduzione è integrale per quello che riguarda il riassunto (v. sopra) e per la prima parte dell'articolo, molto importante per il suo carattere generale. Il resto dell'articolo, piuttosto tecnico e molto lungo, si conclude con i risultati dell'analisi, in parte riportati nel riassunto. Alcune precisazioni e altre osservazioni di particolare interesse sono sintetizzate nel testo seguente.

- La MCID era esplicitamente citata come base della definizione della numerosità del campione solo nell'11% degli articoli analizzabili.
- L'importanza clinica dei risultati era menzionata nel 74% degli articoli; solo nel 50% di essi era oggetto di una discussione esplicita.
- La giustificazione addotta per la definizione della MCID era particolarmente carente: su 4 livelli di validità della giustificazione (decrescenti da 1 a 4) la giustificazione era così distribuita nei 20 articoli analizzabili: 1° livello, nessun articolo; 2° livello, 1 articolo; 3° livello, 4 articoli; 4° livello, 15 articoli.

Commento

L'articolo di Chan et al. riprende un problema che periodicamente (ma non troppo spesso) riappare nella stampa medica: l'importanza clinica dei vantaggi terapeutici "statisticamente significativi" evidenziati nelle sperimentazioni cliniche. La prima parte dell'articolo, di maggior interesse generale, è stata qui tradotta integralmente; della seconda parte sono sintetizzati i punti principali, e l'interpretazione conclusiva. La tesi generale dell'articolo, pienamente condivisa da questo commento, è che significatività statistica e rilevanza clinica non sempre coincidono, che risultati "statisticamente significativi" possono essere clinicamente non importanti e che, viceversa, effetti terapeutici "statisticamente non significativi" possono veicolare il messaggio di una terapia clinicamente utile. Il testo dell'articolo espone fra l'altro una modalità per mettere a confronto la MCID di una terapia valutata in una sperimentazione con il risultato della sperimentazione (espresso come differenza fra la risposta al trattamento sperimentale e quella al trattamento di controllo, con relativo intervallo di confidenza). La modalità riportata nell'articolo è qui illustrata in tre schemi (v. Figura 1); da notare che in tutti e tre gli schemi i risultati degli RCT sono "statisticamente significativi".

Figura 1. Posizione di MCID e intervallo di confidenza



A. La MCID (*) è posta a destra del limite inferiore dell'intervallo di confidenza della differenza fra risultato del trattamento sperimentale e di controllo: è probabile che il risultato del trial sia clinicamente importante.
B. La MCID (*) è posta a sinistra del limite superiore dell'intervallo di confidenza della differenza fra risultato del trattamento sperimentale e di controllo: è probabile che il risultato del trial non sia clinicamente importante.
C. La MCID (*) è all'interno dell'intervallo di confidenza della differenza fra risultato del trattamento sperimentale e di controllo: l'importanza clinica del risultato del trial è dubbia.

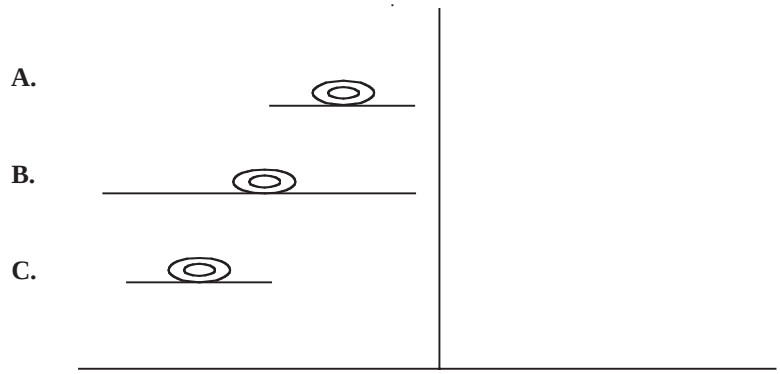
Si è già osservato che il problema dell'interpretazione dei risultati degli RCT dal punto di vista della rilevanza clinica non è nuovo (6,7) e lo stesso termine di MCID era stato introdotto nel 1989 (4). Un contributo che facilita (ma non è sufficiente a consentire) l'interpretazione clinica dei risultati degli RCT è fornito dall'intervallo di confidenza (8), di uso crescente a integrazione o al posto del valore di **p,** e incluso nelle raccomandazioni di CONSORT (1).▲

BOX 1

Alcune considerazioni correlate all'uso dell'intervallo di confidenza

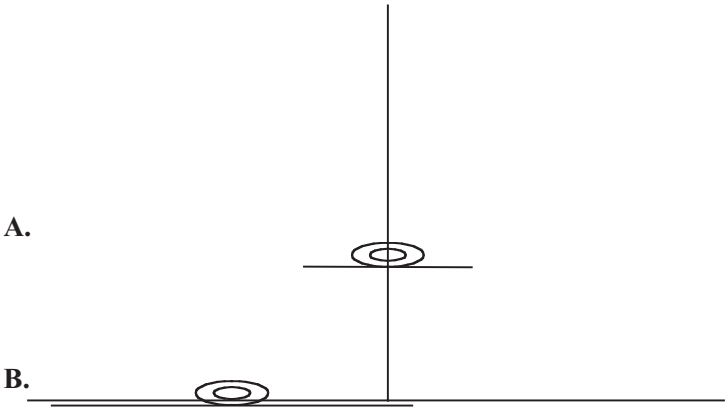
Le Figure 2 e 3, da (8) mod., illustrano come l'uso dell'intervallo di confidenza (IC) fornisce informazioni clinicamente rilevanti non ricavabili dal valore di **p.** Nella Figura 2 sono esaminati tre risultati di RCT, statisticamente significativi ($p < 0,05$), come indica anche il fatto che i rispettivi IC non attraversano la linea di equivalenza. Gli intervalli di confidenza forniscono però altre informazioni. Nello studio A, l'IC ristretto indica che il risultato è riproducibile (e quindi attendibile); l'effetto terapeutico è però modesto (l'IC si estende fino in prossimità della linea di equivalenza). Nello studio B, l'IC è ampio e si estende fino in prossimità della linea di equivalenza: l'effetto terapeutico potrebbe essere modesto o notevole, l'informazione è poco riproducibile. Nello studio C, l'IC ristretto garantisce la riproducibilità del risultato, e la distanza dalla linea di equivalenza indica un effetto terapeutico notevole (e quindi clinicamente rilevante).

Figura 2. Nei tre studi, $p < 0,05$, risultato "statisticamente significativo".
Per le informazioni ulteriori fornite da IC, v. testo.



Nella Figura 3 sono esaminati due risultati di RCT, statisticamente non significativi ($p > 0,05$, IC che attraversa la linea di equivalenza). I tre RCT forniscono però risultati che non sono sovrapponibili. Nel primo studio (A) il trattamento sperimentale e quello di controllo hanno un risultato sovrapponibile (che cade sulla linea di equivalenza), e l' IC è ristretto, indicando che il risultato è riproducibile. Nel secondo studio (B), l'IC ampio indica una scarsa riproducibilità del risultato; inoltre, la parte maggiore dell'IC è a sinistra della linea di equivalenza, non si può escludere che il trattamento sperimentale sia efficace, sono necessari altri studi.

Figura 3. Nei due studi, $p > 0,05$, risultato "statisticamente non significativo".
Per le informazioni ulteriori fornite da IC, v. testo.



BOX 2

In conclusione:

- La rilevanza clinica dei risultati degli RCT dipende criticamente dalla qualità degli *end point*: è ovvio infatti che modificazioni in senso favorevole di un *end point* surrogato possono non essere clinicamente rilevanti, anche se statisticamente significative, se non corrispondono a un reale vantaggio clinico.
- Molte sperimentazioni recenti e attuali includono migliaia (e anche decine di migliaia) di pazienti (*megatrial*). Lo scopo è quello di *reliably detect or refute moderate but still worthwhile treatment effects* (9); esempi di "effetti terapeutici moderati ma ancora clinicamente rilevanti (*worthwhile*)" evidenziati da *megatrials* sono quelli dell'aspirina nei pazienti con sospetto infarto del miocardio, del tamoxifene come terapia adiuvante nel carcinoma della mammella, ed altri (9). D'altra parte, la dimensione dei *megatrial* può rendere "statisticamente significativi" effetti terapeutici assai modesti, di dubbia o anche nulla rilevanza clinica. Guardare alla riduzione assoluta anziché a quella relativa del rischio può fornire un punto di vista più conservativo (e clinicamente più appropriato) dell'efficacia dei trattamenti.
- L'articolo di Chan et al. introduce il bilancio tra efficacia ed eventi avversi dei trattamenti come criterio di valutazione della MCID, e sulla stessa linea si pongono altri studi sopra citati (4,7). Va tenuto presente però che assai spesso gli RCT non riescono a identificare (o a riportare) eventi avversi dei trattamenti in sperimentazione, che si evidenziano poi con l'estensione post-marketing a popolazioni assai più numerose di pazienti non selezionati, con conseguente, e non raro, ritiro dei farmaci in questione (10-14). Sembra dunque evidente che, per una reale valutazione della MCID, gli RCT non sono sufficienti ed è anche necessaria la ricerca di segnalazioni di eventi avversi, rilevati post-marketing.
- L'interpretazione dei risultati degli RCT rigidamente focalizzata su parametri statistici è stata criticata come un esempio di *statistical reductionism*, che tende a sovrapporsi e a prevalere rispetto alle valutazioni in termini di beneficio clinico (15). Il vantaggio dell'interpretazione dei risultati degli RCT basata su parametri statistici è che questi sono matematicamente calcolabili, quantitativi, riproducibili e (forse illusoriamente) obiettivi, mentre il concetto di MCID è intrinsecamente dipendente dal giudizio del medico e pertanto largamente soggettivo. È tuttavia necessario tener presente che i risultati statistici degli RCT riguardano il paziente medio affetto da una malattia e definito da precisi criteri di inclusione ed esclusione; che all'interno di un RCT si trovano pazienti che per le loro caratteristiche di base possono avvantaggiarsi molto, poco, o per nulla di un trattamento statisticamente efficace; che trattamenti statisticamente efficaci possono non ripetersi se l'IC è ampio e si estende fino in prossimità della linea di equivalenza, o se cambiano le caratteristiche dei pazienti in un altro RCT. Queste considerazioni, e altre non sviluppate per brevità, suggeriscono che la medicina attuale deve fare un maggior uso di "*Clinical Judgement*" (16) nell'interpretazione e applicazione degli RCT e, più in generale, della ricerca clinica e deve cercare di migliorare la misurabilità e riproducibilità degli indici (sintomi, segni, qualità di vita) che del *clinical judgement* sono le componenti (17-20).

Bibliografia

1. Moher D et al. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *Ann Intern Med* 2001;**134**:657-62.
2. Sackett DL et al. *Clinical epidemiology: a basic science for clinical medicine*. 2nd ed. Boston: Little, Brown and Company, 1991.
3. Greene WL et al. Claims of equivalence in medical research: are they supported by the evidence? *Ann Intern Med* 2000;**132**:715-22.
4. Jaeschke R et al. Measurement of health status: ascertaining the minimal clinically important difference. *Control Clin Trials* 1989;**10**:407-15.
5. Hart RG et al. Aspirin for the primary prevention of stroke and other major vascular events. *Arch Neurol* 2000;**57**:326-32.
6. Guyatt GH et al for the Evidence-based medicine working group (EBM) Users'guides to the medical literature. IX. A method for grading health care recommendations. *JAMA* 1995;**274**:1800-4.
7. Sinclair JC et al. When should an effective treatment be used? Derivation of the threshold number needed to treat and the minimum event rate for treatment. *J Clin Epidemiol* 2001;**54**:253-62.
8. Borenstein M. The case for confidence intervals in controlled clinical trials. *Control Clin Trials* 1994;**15**:411-28.
9. Collins R et al. Large-scale randomised evidence: trials and overviews. In: Non-random reflections on health services research, Maynard & Chalmers Eds, London: BMJ Publishing Group, 1997:197-230.
10. Ioannidis JPA, Lau J. Completeness of safety reporting in randomized trials. An evaluation of 7 medical areas. *JAMA* 2001;**285**:437-43.
11. Ernst E, Pittler MH. Assessment of therapeutic safety in systematic reviews: literature review. *BMJ* 2001;**323**:546.
12. Friedman MA et al. The safety of newly approved medicines: do recent market removals mean there is a problem? *JAMA* 1999;**281**:1728-34.
13. Brewer T, Colditz GA. Postmarketing surveillance and adverse drug reactions. Current perspectives and future needs. *JAMA* 1999;**281**:824-9.
14. ABC degli studi clinici. Sperimentazione controllata e randomizzata (RCT). *BIF* 2000;**1**:33-8.
15. Feinstein AR. Twentieth Century paradigms that threaten both scientific and humane medicine in the twenty-first century. *J Clin Epidemiol* 1996;**49**:615-7.
16. Feinstein AR. Clinical Judgement. Baltimore: The Williams & Wilkins Co, 1967.
17. Sackett DL. The Rational Clinical Examination: A Primer on the precision and accuracy of the clinical examination. *JAMA* 1992;**267**:2638-44.
18. Redelmeier DA et al. Problems for clinical judgement: 1. Eliciting an insightful history of the present illness. *CMAJ* 2001;**164**:647-51.
19. Redelmeier DA et al. Obtaining a reliable past medical history. *CMAJ* 2001;**164**:809-13.
20. McGee S. Evidence-based Physical Diagnosis. Philadelphia: WB Saunders Co, 2001.