- Copyright - Il Pensiero Scientifico Editore downloaded by IP 52.14.85.76 Thu, 25 Apr 2024, 21:07:27

Effetto statisticamente rilevante ed effetto clinicamente rilevante

Catherine Klersy

G Ital Aritmol Cardiostim 2001;1:23-25

Servizio di Biometria ed Epidemiologia Clinica
Direzione Scientifica
IRCCS Policlinico San Matteo, Pavia

Vi è una certa confusione nella letteratura biomedica, e di conseguenza nella mente del ricercatore, fra effetto *statisticamente* significativo ed effetto *clinicamente* (scientificamente) significativo o rilevante. Non sempre (quasi mai?) negli articoli pubblicati viene discussa l'entità dell'effetto dimostrato in termini di impatto clinico; spesso gli autori si accontentano di riportare la presenza di una significatività statistica. Tuttavia, è importante notare come un effetto statisticamente significativo possa sempre essere messo in evidenza, a condizione che il campione studiato sia sufficientemente ampio. Questo può essere facilmente compreso esaminando la formulazione generale del test statistico:

$$\frac{\text{valore osservato} - \text{valore atteso}}{\text{variabilità del campione}}$$

Il valore del test sarà tanto più elevato (e quindi il p-value tanto più piccolo) quanto più piccolo il denominatore (scarsa variabilità ottenibile per elevata numerosità campionaria) e/o tanto più grande il numeratore (effetto valutato). Quindi otterremo test statistici significativi se l'effetto che stiamo valutando è importante, ma anche, in presenza di un campione numeroso, se non lo è.

Significatività statistica

La significatività statistica è un concetto probabilistico: rappresenta la probabilità di ottenere un falso positivo quando si afferma che esiste un certo effetto. Per convenzione, questa probabilità deve essere inferiore al 20% ($p < 0,05$) per affermare che un effetto è statisticamente significativo. È da sottolineare, tuttavia, che una divisione arbitraria dei risultati in "significativi" o "non significativi" in base al valore di p non era nelle intenzioni dei fondatori dell'inferenza statistica. Il p-value, o livello di significatività, misura l'entità dell'evidenza contro l'ipotesi nulla (di non effetto). Tanto più piccolo è il valore di p, tanto maggiore è l'evidenza contro l'ipotesi nulla. Un valore di $p = 0,05$ non rappresenta un'evidenza importante, mentre si può ragionevolmente affermare che $p < 0,001$ deponga per una forte evidenza a favore del rifiuto dell'ipotesi nulla. È opportuno, pertanto, che nei risultati di un lavoro

- Copyright - Il Pensiero Scientifico Editore downloaded by IP 52.14.85.76 Thu, 25 Apr 2024, 21:07:27

venga presentato il valore di p calcolato, senza riferimento a una soglia arbitraria. Va sottolineato come la letteratura medica attuale sia dominata da questa divisione significativo/non significativo. Questo ha due conseguenze principali: differenze clinicamente importanti osservate in studi piccoli sono dichiarate non significative e quindi ignorate; tutti i risultati significativi vengono considerati espressione di un effetto reale, sia questo legato a un trattamento oppure a un fattore di rischio.[1]

Effetto clinicamente rilevante

Ben diverso è il concetto di effetto clinicamente rilevante. L'effetto che vogliamo misurare può essere una differenza di medie, una differenza di proporzioni, un rischio relativo, un rischio assoluto, un numero di casi da trattare per prevenire un evento, ecc. a seconda del problema oggetto dello studio. In altri termini, non è sufficiente dire, per esempio, che un trattamento funziona, è necessario dire anche di *quanto* funziona.[1,2,3] Questo significa formulare non solo un'ipotesi nulla di non effetto, ma anche un'ipotesi alternativa per quantizzare l'effetto atteso nella progettazione di uno studio. La dimostrazione di un effetto clinicamente rilevante deve essere l'obiettivo primario di ogni studio. Un effetto clinicamente rilevante potrebbe essere definito come il minimo beneficio ottenibile, tale da indurre il medico a cambiare strategia nei confronti del paziente, per esempio consigliando un nuovo trattamento o verificando la presenza di un certo fattore di rischio per correggerlo.[4]

Interpretazione dei risultati di uno studio

Da questi due concetti di significatività statistica e rilevanza clinica deriva che l'assenza di evidenza non è evidenza di assenza:[2] un risultato non significativo non è sinonimo di studio negativo o di non effetto; tutto quello che si può dire è che manca l'evidenza di un effetto. Una maggiore comprensione del fenomeno viene facilitata riportando la stima dell'effetto e il suo intervallo di confidenza, per esempio al 95%;[5] questo è l'intervallo in cui sarà compresa la stima 95 volte su 100. In questo modo si potrà apprezzare la rilevanza clinica dell'effetto in base al valore della sua stima e a quello dei limiti superiore e inferiore dell'intervallo. L'ampiezza dell'intervallo potrà suggerire se la mancata significatività statistica è prevalentemente legata a un difetto di potenza legato a un'insufficiente numerosità campionaria (intervallo ampio). È implicito il fatto che la validità di queste stime e la loro generalizzabilità alla popolazione sottostante sono strettamente dipendenti da un disegno dello studio volto a minimizzare i bias.

Numerosità campionaria

La quantizzazione dell'effetto clinicamente rilevante è particolarmente importante nella programmazione di uno studio clinico: su di esso si basa la definizione della numerosità campionaria necessaria per dimostrare l'end-point primario. Illustriamo un esempio di due gruppi di trattamento in cui viene misurato il tasso di eventi; gli elementi necessari per il calcolo della numerosità campionaria sono:

- l'errore α (falsi positivi): per convenzione fissato a 0,05 o 0,01;
- l'errore β (falsi negativi) o la potenza 1-β: per convenzione fissata a 0,80 o 0,90;
- la specificazione delle code del test (1 o 2);
- il tasso di eventi attesi nel gruppo di controllo: desunto dalla letteratura e dall'esperienza clinica;
- l'effetto clinicamente rilevante o minima differenza importante.

L'entità dell'effetto clinicamente rilevante può avere effetti drammatici sulla numerosità campionaria programmata: piccoli effetti portano a grandi dimensioni campionarie e viceversa. Nella Tabella I si può vedere come, a parità di errore α, di potenza e di proporzione di eventi nel gruppo di controllo, il numero di casi per gruppo da arruolare vari ampiamente con il variare della proporzione di eventi del gruppo trattato, cioè dell'effetto giudicato clinicamente rilevante. Per il computo è stato utilizzato nQuery Advisor 4.0 (Statistical Solutions, Cork, Irlanda).

La determinazione dell'effetto clinicamente rilevante è arbitraria e di responsabilità del ricercatore. Non

- Copyright - Il Pensiero Scientifico Editore downloaded by IP 52.14.85.76 Thu, 25 Apr 2024, 21:07:27

TABELLA I Test χ^2 per l'uguaglianza delle proporzioni (2 gruppi) (ipotesi nulla odds ratio = 1; uguale numerosità dei gruppi)

Effetto clinicamente rilevante	Riduzione del 50%	Riduzione del 33%	Riduzione del 20%
Livello di significatività,	0,050	0,050	0,050
α 1 o 2 code?	2	2	2
Gruppo 1, proporzione, P1	0,600	0,600	0,600
Gruppo 2, proporzione, P2	0,300	0,400	0,480
Odds ratio	3,496	2,252	1,626
Potenza (%)	80	80	80
N. per gruppo	42	97	270

esistono regole scritte; nella pratica clinica può essere influenzata da diversi fattori quali: 1) gli effetti collaterali della terapia e i suoi costi; 2) l'importanza dell'esito che il trattamento sta cercando di prevenire; 3) l'esperienza clinica del medico, in particolare quando deve valutare l'effetto di un fattore di rischio; 4) come è misurato l'effetto clinicamente rilevante (medie, proporzioni, rischio relativo, rischio assoluto, ecc.). Un recente lavoro pubblicato da van Walraven et al.[4] illustra l'effetto di questi parametri sul giudizio della minima differenza importante; evidenzia come la dimensione campionaria calcolata possa variare da un centinaio di casi a più di 3000.

È fondamentale, pertanto, che questo effetto venga valutato in maniera attenta dai ricercatori, possibilmente confrontando le proprie esperienze e rapportandosi alla letteratura esistente sul fattore di rischio/trattamento da valutare o su fattori di rischio/trattamenti simili, per: 1) non disegnare studi sotto- o sovradimensionati; 2) poter dare un giudizio dal punto di vista clinico e non solo statistico (paradossalmente, per un medico) sui risultati ottenuti nei propri studi o negli studi riportati in letteratura.

Conclusioni

I risultati della ricerca medica non andrebbero riportati come significativi o non significativi, ma interpretati nel contesto dello studio e di altre evidenze a disposizione. A tal fine l'entità dell'effetto e la rilevanza clinica dei relativi limiti di confidenza vanno commentate. Va discusso se la direzione di un effetto è stata stabilita con una ragionevole certezza e se l'entità di questo effetto è tale da essere di rilevanza clinica e/o in termini di salute pubblica.

Bibliografia

1. Sterne JAC, Davey Smith GD: Sifting the evidence–what's wrong with significance tests? *BMJ* 2001;322:226-231.
2. Altman DA, Bland MJ: Absence of evidence is not evidence of absence. *BMJ* 1995;311:485.
3. Tarnow-Mordi WO, Healy MJR: Distinguishing between "no evidence of effect" and "evidence of no effect" in randomized clinical trials and other comparisons. *Arch Dis Chil* 1999;80:210-211.
4. van Walraven C, Mahon JL, Moher D, et al.: Surveying physicians to determine the minimal important difference: implications for sample size calculation. *J Clin Epidemiol* 1999;52:717-723.
5. Cox DR: Another comment on the role of statistical methods. *BMJ* 2001;322:231.

Indirizzo per la corrispondenza

Catherine Klersy
Servizio di Biometria ed Epidemiologia Clinica
IRCCS Policlinico S. Matteo
27100 Pavia
e-mail: klersy@smatteo.pv.it