

PILLOLE DI METODOLOGIA DELLA RICERCA

Analisi per sottogruppi nei trial clinici

Fidarsi è bene... non fidarsi è meglio

I trial controllati e randomizzati - *Randomized Controlled Trials* (RCTs) - costituiscono il gold-standard per dimostrare l'efficacia degli interventi sanitari. Considerato che richiedono consistenti investimenti degli sponsor (pubblici e/o privati), oltre che di ricercatori e partecipanti, nei trial vengono effettuate numerose analisi statistiche alla ricerca di qualche differenza statisticamente significativa tra i due gruppi di partecipanti. Oltre alla possibilità di analizzare l'efficacia dell'intervento in studio su outcome secondari, le analisi per sottogruppi (AS) identificano nella popolazione in studio un certo numero di sottogruppi per verificare se l'outcome di interesse si presenta più frequentemente in qualcuno di essi rispetto ad altri: ad es. maggior effetto di un nuovo farmaco nelle donne rispetto agli uomini, nei diabetici rispetto ai non diabetici, etc.

Considerato che i documenti regolatori sulle sperimentazioni cliniche (in particolare sui farmaci) incoraggiano l'utilizzo appropriato delle AS - che possono anche modificare le raccomandazioni cliniche delle linee guida - secondo recenti survey, oltre il 70% dei RCTs effettua AS. Tuttavia, il legittimo desiderio dei ricercatori di identificare eterogeneità nella risposta terapeutica nella popolazione arruolata in un trial, contrasta con gli strumenti metodologici e statistici che spesso producono risultati distorti. Per tali ragioni, il CONSORT 2010 - ultima versione delle linee guida per il reporting dei RCTs - rispetto alle AS, raccomanda di descrivere adeguatamente:

- Le metodologie utilizzate per analisi aggiuntive, come le analisi per sottogruppi e le analisi "aggiustate" (item 12b)
- I risultati di tutte le analisi effettuate, incluse le analisi per sottogruppi e le analisi "aggiustate", distinguendo quelle pre-definite da quelle esplorative (item 18).

1. Problemi statistici delle analisi per sottogruppi

Sono conseguenti al rischio di risultati sia falsamente positivi, sia falsamente negativi.

1.1. Analisi multiple e risultati falsamente positivi

Effettuare un'analisi statistica fissando a 0.05 il livello di significatività, equivale ad accettare che la differenza tra due trattamenti nel 5% dei casi non esiste (errore alfa = falso positivo). Considerato che in un RCT possono essere definiti innumerevoli sottogruppi (genere, età, razza, centro di arruolamento, fumo, stadio di malattia, patologie concomitanti, etc), bisogna tenere presente che all'aumento delle AS effettuate cresce la probabilità che almeno una sia statisticamente significativa, ma falsamente positiva (tabella).

N° di analisi per sottogruppi	Probabilità di risultato falsamente positivo
1	5%
2	10%
5	23%
10	40%
20	64%
50	92%
100	99%

Uno degli esempi più clamorosi di come i risultati di AS possano essere condizionati solo da fattori casuali è rappresentato dallo studio ISIS-2: nei pazienti con infarto miocardico acuto (IMA) l'aspirina riduce la mortalità a 5 settimane in tutti i pazienti tranne... per i nati sotto il segno della Bilancia o dei Gemelli. In tal caso, l'assenza di plausibilità biologica non ha modificato le raccomandazioni cliniche delle linee guida per i pazienti con IMA.

1.2. Potenza statistica e risultati falsamente negativi

La maggiore parte dei trial arruolano un numero di partecipanti sufficiente per testare l'ipotesi primaria. Di conseguenza le analisi statistiche effettuate sui sottogruppi hanno una potenza sufficiente solo per identificare notevoli effetti dell'intervento sanitario. La ridotta compliance dei pazienti e gli aggiustamenti richiesti dai test multipli riducono ulteriormente la potenza, per cui:

- l'effetto del trattamento su numerosi sottogruppi può non raggiungere la significatività statistica, magari ottenuta sull'intero campione della popolazione arruolata;
- reali differenze nella risposta all'intervento sanitario (eterogeneità), potrebbero non essere identificate nei sottogruppi per insufficiente numerosità del campione.

Pertanto, se da un lato le AS non hanno una sufficiente potenza statistica per mettere in evidenza reali differenze tra i sottogruppi (errore beta = falsi negativi), quando utilizzate ripetutamente per rilevare differenze tra numerosi sottogruppi, identificano effetti spuri (errore alfa = falsi positivi). Per riconciliare la necessità di identificare vere differenze tra i sottogruppi, minimizzando il rischio di rilevare e pubblicare risultati falsamente positivi è necessario:

- accettare che i risultati delle AS, per definizione, sono ipotesi da testare ulteriormente in studi successivi;
- seguire alcune indicazioni pratiche in fase di disegno del trial, di analisi, reporting e interpretazione.

2. Disegno del trial

Appropriatezza dei sottogruppi. I sottogruppi basati su caratteristiche identificate dopo la randomizzazione dovrebbero essere evitati, perchè fortemente influenzate dall'intervento assegnato. Ad esempio, pazienti *responders vs non-responders* o pazienti complianti vs non-complianti al trattamento rappresentano classici esempi di "sottogruppi inopportuni".

Definizione a priori dei sottogruppi. In generale, le AS dovrebbero essere definite a priori (*pre-planned*), sulla base di meccanismi biologici conosciuti o in relazione a dati di studi precedenti. Idealmente, nel protocollo dovrebbero essere descritti sia la plausibilità biologica di ogni singolo sottogruppo, sia la direzione attesa degli effetti del trattamento. Viceversa, i sottogruppi identificati dopo la conclusione dei trial e la conseguente analisi del dataset dovrebbero essere guardate con un certo scetticismo. Occupano una posizione intermedia, i sottogruppi non specificati nel protocollo, ma identificati nel corso del trial in seguito alla migliore risposta terapeutica in uno specifico sottogruppo.

3. Analisi e reporting

I ricercatori sono tenuti a riportare tutte le informazioni necessarie per consentire al lettore di valutare la validità delle AS riportate. In particolare:

- Dichiarare il numero totale delle AS effettuate e di quelle riportate, per fornire una stima indiretta della probabilità di risultati falsamente positivi (tabella).
- Descrivere per ciascun sottogruppo la plausibilità biologica e/o il riferimento a sottogruppi identificati da studi precedenti.
- Per ciascuna AS riportare se è stata definita nel protocollo (*pre-planned*), se identificata nel corso del trial, oppure effettuata a posteriori dopo la conclusione dello studio (*post-hoc*).
- Fornire per ciascun sottogruppo, i numeri assoluti: soggetti ed eventi rilevati per ciascun braccio del trial.

4. Interpretazione

Considerato che, rispetto all'intera popolazione, le AS hanno una potenza limitata nell'identificare l'effetto terapeutico, nel report del trial - in particolare nell'abstract e nelle conclusioni - dovrebbe essere posta enfasi sul risultato complessivo. Infatti, considerato l'aumento della probabilità di risultati falsamente positivi quando vengono effettuate numerose AS, non sorprende il fatto che numerosi trial negativi riportino effetti statisticamente significativi in uno o più sottogruppi.

In conclusione, è saggio presumere che i risultati ottenuti da un trial su tutta la popolazione possono essere applicati a tutti, tranne se *good evidence* dimostrino il contrario.

Nino Cartabellotta
Direttore Scientifico GIMBE

KEY POINTS

- **Nei trial clinici le analisi per sottogruppi vengono effettuate molto frequentemente**
- **All'aumentare del numero delle analisi per sottogruppi effettuate cresce il rischio di risultati falsamente positivi**
- **La limitata potenza statistica delle analisi per sottogruppi determina spesso risultati falsamente negativi**
- **Le analisi per sottogruppi definite *post-hoc* a scopo esplorativo forniscono evidenze molto deboli**
- **Il CONSORT 2010 definisce tutti i dettagli che i ricercatori devono riportare su metodi e risultati delle analisi per sottogruppi**
- **La verosimiglianza di un "effetto sottogruppo" si articola in un *continuum* da altamente plausibile a estremamente improbabile**
- **La credibilità delle analisi per sottogruppi dipende da: plausibilità biologica, definizione a priori dei sottogruppi, conferma dell'effetto in studi successivi**

Per saperne di più

Letture introduttive

- Sun X, Briel M, Walter SD, et al. Is a subgroup effect believable? Updating criteria to evaluate the credibility of subgroup analyses. *BMJ* 2010;340:c117.
- Wang R, Lagakos SW, Ware JH, et al. Statistics in medicine--reporting of subgroup analyses in clinical trials. *N Engl J Med* 2007;357:2189-94.
- Fletcher J. Subgroup analyses: how to avoid being misled. *BMJ* 2007;335:96-7.
- Cook DI, Gebski VJ, Keech AC. Subgroup analysis in clinical trials. *Med J Aust* 2004; 180: 289-291.

Approfondimenti

- Schulz KF, Grimes DA. Multiplicity in randomised trials II: subgroup and interim analyses. *Lancet* 2005;365:1657-61.
- Rothwell PM. Subgroup analysis in randomised controlled trials: importance, indications, and interpretation. *Lancet* 2005;365:176-86.
- Simes RJ, Gebski VJ, Keech AC. Subgroup analysis: application to individual patient decisions. *Med J Aust* 2004;180:467-9.
- Brookes ST, Whitely E, Egger M, et al. Subgroup analyses in randomized trials: risks of subgroup-specific analyses; power and sample size for the interaction test. *J Clin Epidemiol* 2004;57:229-36.
- Assmann SF, Pocock SJ, Enos LE, et al. Subgroup analysis and other (mis)uses of baseline data in clinical trials. *Lancet* 2000;355:1064-9.
- Horton R. From star signs to trial guidelines. *Lancet* 2000;355:1033-4.