

zienti, mentre, nel secondo caso, fissare cicli di terapia di durata inferiore a 28 giorni avrebbe richiesto una più frequente somministrazione dei questionari.

Confronto. Mediante gli NCI-CTC, nello studio A è stata considerata la presenza (o meno) di fatigue (di qualunque livello) e di fatigue severa in tutti i pazienti randomizzati ai due trattamenti, nell'intero periodo di tempo dello studio, e si è visto che non vi è alcuna differenza.

Nello studio B, invece, sono stati selezionati i pazienti con maxF e intF non inferiori a 0,5 e solo per questi è stato valutato l'andamento delle fatigue nel tempo: **non sono stati pertanto considerati gli altri pazienti.** Nel confronto tra i risultati ottenuti occorre tener conto che potrebbero essersi presentati vari inconvenienti, tra cui uno particolarmente grave: i pazienti non considerati (quelli che presentavano un cut-off inferiore a 0,5) potrebbero aver avuto un incremento di fatigue, nei periodi successivi al basale, in misura più marcata nel gruppo di trattamento con abiraterone rispetto al gruppo di controllo; se così fosse accaduto, tale evento sarebbe sfuggito all'analisi riportata nello studio B. D'altronde, se assumiamo che i due metodi di valutazioni (NCI-CTC e BFI) diano luogo a risultati abbastanza concordanti, la situazione descritta potrebbe essere vicina a quella reale, in quanto, in tal modo, si ricomporrebbe il quadro esposto nello studio A (nessuna differenza tra abiraterone e placebo): a compensare la maggiore efficacia di abiraterone nei pazienti che avevano una almeno discreta fatigue, ci sarebbe la presenza di una maggiore insorgenza di fatigue o di un suo peggioramento, rispettivamente in coloro che non l'avevano o la percepivano lieve nel gruppo dei soggetti trattati con abiraterone. Le conclusioni dello studio B, pertanto, potrebbero essere parziali ed irrealistiche.

In secondo luogo, data l'arbitrarietà della scelta del cut-off, nessuno può garantire che i pazienti con livelli di fatigue immediatamente inferiori a quelli stabiliti dal cut-off (ad es., 0,45) si comportino in modo analogo a quello dei pazienti osservati.

In conclusione, a nostro avviso, l'effetto specifico di abiraterone sulla fatigue non appare sufficientemente provato e, quindi, non resta che accogliere il verdetto di parità tra abiraterone e placebo sancito nello studio A. •

Bibliografia

1. Mendoza TR, Wang XS, Cleeland CS, et al. The rapid assessment of fatigue severity in cancer patients: use of the Brief Fatigue Inventory. *Cancer* 1999; 85: 1186-96.
2. Holzner B, Kemmler G, Sperner-Unterwieser B, et al. Quality of life measurement in oncology: a matter of the assessment instrument? *Eur J Cancer* 2001; 37: 2349-56.
3. Minton O, Stone P. A systematic review of the scales used for the measurement of cancer-related fatigue (CRF). *Ann Oncol* 2009; 20: 17-25.
4. Apolone G, Ballatori E, Mosconi P, Roila F. *Misurare la qualità di vita in Oncologia. Roma: Il Pensiero Scientifico Editore, 1997; pp. XII + 112.*

Statistica per concetti

Randomizzazione

Nel lavoro sintetizzato nella scheda 2 compare un uso per certi versi non comune della randomizzazione. Pertanto, si è deciso di dedicare a questo argomento la presente rubrica, discutendo, alla fine, le peculiarità della randomizzazione adoperata.

La randomizzazione, operazione fondamentale nelle scienze sperimentali, può definirsi come l'allocazione **rigorosamente casuale** delle unità sperimentali ai gruppi di trattamento. L'avverbio "rigorosamente" indica che l'assegnazione casuale delle unità sperimentali ai trattamenti è eseguita in base ad un metodo ben preciso, e non con il significato di "senza un criterio", che l'aggettivo "casuale" ha spesso nel linguaggio comune.

Ad esempio, dovendo randomizzare 16 pazienti a due terapie, A e B, si può ricorrere alle tavole di numeri aleatori, costruite con una procedura di estrazione casuale di numeri (gioco del lotto): è sufficiente immaginare una fila di 16 pazienti, leggere 8 numeri di due cifre ciascuno compresi tra 01 e 16, e decidere di assegnare i corrispondenti pazienti al trattamento A; per differenza gli altri verranno assegnati a B. Ad es., se i primi 8 numeri letti sono 09, 04, 12, 01, 05, 14, 16, 08, i pazienti che, nell'ordine della fila occuperanno i posti 1, 4, 5, 8, 9, 12, 14, 16 verranno assegnati ad A, gli altri a B.

Qualunque sia il metodo seguito (tavole aleatorie, generazione di numeri pseudo-casuali mediante computer), la randomizzazione deve essere sempre

- **riproducibile**, nel senso che, una volta descritto il procedimento, chiunque lo applichi ottenga sempre lo stesso risultato (quindi, il lancio di una moneta non può essere utilizzato per randomizzare);
- **imprevedibile**, cioè che non sia quasi mai possibile prevedere a quale trattamento verrà assegnata la prossima unità. Infatti, se il medico sperimentatore sapesse a quale trattamento verrà assegnato il successivo paziente, potrebbe essere condizionato nella decisione di reclutarlo (ad es., se sa che il paziente sarà randomizzato al trattamento più tossico, potrebbe non proporre l'ingresso nello studio ad un paziente defedato, malgrado egli soddisfi i criteri di selezione).

Gli scopi più importanti della randomizzazione sono:

- a. **produrre un accettabile bilanciamento dei gruppi di trattamento rispetto ai fattori prognostici noti e sconosciuti;**
- b. **fornire una seconda base logica al test statistico** che verrà usato per la valutazione di efficacia/tollerabilità differenziali.

Per quanto concerne il punto (a), la randomizzazione non garantisce il perfetto bilanciamento dei gruppi sperimentali in relazione ai fattori prognostici (anzi, è quasi impossibile che lo ottenga), ma assicura che, con alta probabilità, non si verifichi uno sbilanciamento tale da poter attribuire ad esso, e non alla diversa efficacia/tollerabilità dei trattamenti, il risultato significativo conseguito.

Ad esempio, 400 pazienti siano randomizzati a due bracci di trattamento: 200 ad A e 200 a B. La presenza di diabete mellito sia un importante fattore prognostico e siano 80, tra i 400, i pazienti diabetici. Se ce ne fossero 40 in un braccio e 40 nell'altro, il diabete non avrebbe alcuna influenza sul risultato. Randomizzando i pazienti, questa parità è un evento quasi impossibile da ottenere, ma la probabilità che si verifichi uno sbilanciamento tale da poter attribuire ad esso, e non alla diversa efficacia/tollerabilità, dei trattamenti il risultato osservato è nulla (ad es., meno di 30 diabetici in un gruppo e, quindi, più di 50 nell'altro ha probabilità pari a 0 di verificarsi).

Per quanto riguarda l'obiettivo (b), vi è da premettere che l'inferenza statistica normalmente usata per l'analisi dei risultati, è basata sul modello di popolazione, cioè sul concetto che i gruppi a confronto siano campioni estratti da popolazioni. Nella ricerca clinica, l'unico (debole) elemento che possa far riguardare i gruppi sperimentali come campioni è la *consecutività* dell'arruolamento. Quando i bracci di trattamento non possono essere considerati campioni casuali, a giustificare la validità dei risultati ottenuti interviene il principio di randomizzazione che, in prima approssimazione, può essere così schematizzato.

La risposta (ossia ciò che si osserva su ogni paziente al termine dello studio) dipende dal trattamento e dal paziente.

Sotto l'ipotesi nulla di uguale efficacia dei trattamenti (ossia, se i trattamenti sono ugualmente efficaci), la risposta dipende solo dal paziente. Il paziente viene allora randomizzato al trattamento con la sua risposta.

Se fosse vera l'ipotesi nulla, non sarebbe da attendersi uno squilibrio sostanziale tra i due gruppi rispetto alla risposta, perché, essendo stati i pazienti

randomizzati, un forte squilibrio ha probabilità nulla di presentarsi (v. quanto esposto nel punto (a)). Quindi, se si verifica uno squilibrio importante, non resta che respingere l'ipotesi nulla al livello di significatività precisato e riconoscere ai trattamenti una diversa efficacia.

La randomizzazione fu introdotta nel campo delle scienze agrarie negli anni '20 del secolo scorso e, successivamente, fu estesa ad altri settori di ricerca. In medicina fu applicata alla fine degli anni '40 e, prima di allora, ogni studio clinico comparativo prestava il fianco a serie obiezioni. Ad esempio, volendo confutare l'idea che il salasso – pratica allora correntemente usata – potesse essere utile nella cura della polmonite, il grande clinico francese Pierre Alexandre Louis nel 1835 pubblicò un lavoro in cui un rilevante numero di pazienti affetti da polmonite furono divisi in due gruppi: tutti furono esposti alle stesse condizioni ambientali (dieta, condizioni igieniche, arieggiamento delle camere di degenza), ma una metà venne curata con salasso, che fu negato all'altra metà. Avendo osservato un all'incirca uguale numero di morti nei due gruppi, al termine dell'articolo, egli concluse che "le polmoniti guariscono da sole, né più né meno di quelle salassate". La critica decisiva mossa a tale studio fu che i due gruppi non erano rassomiglianti per tutte le caratteristiche atte ad influenzare il decesso, per cui poteva essere accaduto che il gruppo dei salassati fosse stato in peggiori condizioni dell'altro e, quindi, che solo l'efficacia del salasso avesse consentito di ottenere una simile mortalità. Ragionamento inoppugnabile, senza randomizzazione: il salasso è restato nella pratica clinica fino alla fine dell'800.

Per concludere, commentiamo i dettagli della randomizzazione così com'è stata usata nel lavoro riportato nella scheda riepilogativa.

Rapporto di randomizzazione 2:1.

Di norma, è conveniente che i due gruppi siano ugualmente numerosi perché, a parità di tutte le altre condizioni, la varianza dello stimatore della statistica test è minore, e, quindi, è più facile accorgersi della diversa efficacia dei trattamenti (cioè, è più probabile che il test risulti significativo se i trattamenti sono diversamente efficaci).

Si sceglie una soluzione diversa (ad es., randomizzazione 2 a 1) se il nuovo farmaco ha dato prove consistenti del suo valore in precedenti studi di fase 2, così da privare, per motivi etici, un minor numero di pazienti del trattamento

potenzialmente più utile. Ovviamente, tale scelta si paga con una maggiore numerosità dei pazienti in studio.

Randomizzazione stratificata.

Quando esistono alcuni fattori (possibilmente pochissimi) che sono fortemente correlati con la risposta, i pazienti possono essere prima classificati rispetto a tali fattori e poi, in ogni gruppo così formato, randomizzati ai trattamenti. In tal modo, rispetto ai fattori di stratificazione, il bilanciamento dei gruppi è perfetto, e non solo approssimativo come invece avviene con la sola operazione di randomizzazione. Nello studio, sono stati considerati 4 fattori di stratificazione, ciascuno con 2 livelli. In totale, quindi, sono stati costituiti 16 gruppi di pazienti, e la randomizzazione è avvenuta all'interno di ciascuno di essi.

La raccomandazione è **non eseguire la randomizzazione stratificata**, o, se proprio la si ritiene necessaria, considerare uno o due fattori, non solo perché la complessità è complice dell'errore, ma anche per evitare di imporre vincoli ad un'operazione che per sua natura dovrebbe essere lasciata del tutto libera nel conseguimento dei suoi obiettivi.

Randomizzazione a blocchi.

Un blocco di randomizzazione è un insieme di indicazioni circa la randomizzazione di un relativamente piccolo numero di pazienti. Ad esempio, se la randomizzazione avviene a blocchi di 12, essendo 2:1 il rapporto di assegnazione dei pazienti ai due trattamenti, ogni blocco di randomizzazione indica quali siano (dei 12) gli 8 pazienti destinati a ricevere il trattamento attivo e quindi quali i 4 che saranno trattati con placebo. In tal modo i due trattamenti sono perfettamente bilanciati, nel loro rapporto, ogni 12 pazienti arruolati.

Negli studi multicentrici è necessario procedere con la costruzione di blocchi di randomizzazione, per fare in modo che in ogni centro sia all'incirca nello stesso rapporto il numero dei pazienti che ricevono i trattamenti a confronto. È però indispensabile che i blocchi non siano eccessivamente piccoli per impedire il più possibile la prevedibilità della randomizzazione.

Enzo Ballatori