- Copyright - Il Pensiero Scientifico Editore downloaded by IP 3.149.243.32 Mon, 29 Apr 2024, 01:09:41

Trattamenti della fatigue... e dintorni

Enzo Ballatori
Statistico medico, Spinetoli

Fausto Roila
SC di Oncologia Medica Azienda Ospedaliera "S. Maria", Terni

Riassunto

Sono stati discussi aspetti positivi e negativi del lavoro sintetizzato nella scheda, sottolineando i problemi che discendono dalla scelta degli strumenti statistici con cui sono stati elaborati i dati.

Parole chiave. *Fatigue, esperimenti pienamente fattoriali, doppio cieco, scale di misurazione, qualità della vita.*

Summary

Treatment of fatigue and other symptoms

Positive and negative aspects of the study summarized in the "scheda" were discussed, highlightening the problems arising from the choice of the statistical tools used to analyze the results.

Key words. *Fatigue, fully factorial experiments, double-blind, measurement scales, quality of life.*

Luci abbaglianti ed ombre scure, oltre a diverse tonalità di grigi, sono presenti nel lavoro sintetizzato nella scheda.

Cominciamo dal titolo in cui sembra sminuita l'importanza del lavoro parlando di "studio di fase II", quando

a. si tratta di uno studio randomizzato controllato con placebo,
b. la risposta valutata è in termini di efficacia e non di attività.

Anzitutto nel testo non sono state precisate due cose fondamentali per valutare la qualità del lavoro: la consecutività del reclutamento dei pazienti nello studio e l'adozione di tecniche di cecità. Il problema è sapere se tali precisazioni mancano perché non sono state adottate o per trascuratezza nella stesura dell'articolo. Il secondo caso sarebbe poco grave, mentre il primo aprirebbe la strada a forti dubbi sulla validità dei risultati. Com'è noto, infatti, la mancanza della consecutività di arruolamento dei pazienti nello studio impedisce di considerare i gruppi sperimentali come campioni casuali, rendendo più deboli le conclusioni raggiunte con i test statistici. La mancata adozione di tecniche di cecità, invece, renderebbe inattendibili i risultati perché, in caso di valutazioni psicometriche, la conoscenza del trattamento cui si è sottoposti può influire pesantemente sulla risposta. Riteniamo che, almeno per l'adozione di tecniche di cecità, lo studio sia stato eseguito in doppio-cieco perché ci sembra altrimenti inverosimile un confronto con placebo. Comunque, in assenza di conoscenze certe, non resta che sospendere il giudizio, tacitamente rimproverando i tre referee del JCO, di norma assai meticolosi, che non hanno richiesto così importanti precisazioni.

Sebbene non sia stato esplicitamente scritto nel lavoro, gli autori hanno adottato un disegno fattoriale completo, assai raro in medicina. Infatti, dopo aver individuato i due fattori sperimentali, quello farmacologico (metilfenidato, MP, vs placebo, PL) e l'intervento psicologico (telefonata terapeutica, TT, vs telefonata di controllo, TC), i pazienti sono stati randomizzati ai 4 trattamenti definiti dalle combinazioni dei livelli dei due fattori sperimentali considerati. La caratteristica fondamentale di un disegno fattoriale completo è che, oltre agli effetti principali dei fattori considerati, è possibile anche valutare le **interazioni**, cioè, nel nostro caso se l'effetto farmacologico del metilfenidato sia potenziato o meno dalla telefonata terapeutica (v. in questo numero Statistica per concetti 1).

La scelta di un disegno fattoriale completo può essere considerata una luce abbagliante del lavoro, ma su di essa si addensano ombre scure.

Ci si sarà chiesto perché tale disegno sia così poco frequente in medicina, malgrado possa fornire conoscenze di notevole importanza per la pratica clinica. La principale ragione è che l'adozione del disegno fattoriale completo comporta una grande dimensione dello studio, spesso incompatibile con i mezzi disponibili. Questa osservazione introduce a quello che sembra il più importante limite dello studio in oggetto: la scarsa numerosità campionaria (al riguardo si osservi che la potenza dello studio non è stata precisata: nell'articolo si parla di "potenza adeguata", che ci siamo permessi di virgolettare anche nella scheda). Della questione si fanno carico (ma solo in parte) gli autori nella discussione: "There is the possibility that the study was underpowered for the reported outcomes because of lower than planned accrual (190 patients instead of the planned 215)". Ma non è certo questo il problema, bensì il fatto che 215 pazienti sarebbero stati comunque pochi per un disegno fattoriale completo. Forse questa è la ragione per cui hanno definito come endpoint primario il solo effetto farmacologico del metilfenidato. Però, se l'interazione non avesse avuto interesse, gli autori avrebbero potuto adottare un disegno a 3 bracci (anziché a 4): controllo (TC + PL), MP, TT. In tal modo si sarebbero ottenute informazioni sull'efficacia del metilfenidato e della telefonata terapeutica confrontando entrambi i trattamenti con quello di controllo, basando però le conclusioni su dimensioni ben più ampie di pazienti.

- Copyright - Il Pensiero Scientifico Editore downloaded by IP 3.149.243.32 Mon, 29 Apr 2024, 01:09:41

SCHEDA

Bruera E, Yennurajalingam S, Palmer JL, et al. Methylphenidate and/or a nursing telephone intervention for fatigue in patients with advanced cancer: a randomized, placebo-controlled, Phase II Trial. J Clin Oncol 2013; 31: 2421-7.

Pazienti con tumore metastatico, con uno score della fatigue ≥ 4 sulla Edmonton Symptom Assessment Scale (ESAS, una scala da 0 a 10) ed uno score normale (≥ 24 su 30) del Mini-Mental State Examination furono arruolati da unità di cure palliative e tra pazienti seguiti da cliniche oncologiche del Texas.
Furono considerati due fattori sperimentali, ciascuno con due modalità:
- metilfenidato (methylphenidate, MP, capsule di 5 mg per os ogni due ore, fino a un massimo di 20 mg/die per 14 giorni) o placebo (PL),
- una telefonata terapeutica (TT) eseguita in forma standardizzata – cioè seguendo un temario – da un infermiere specializzato (da 4 a 6 volte nei 14 giorni di durata dello studio), TT (Telefonata Terapeutica), o una telefonata libera eseguita da un soggetto che non era un infermiere (sempre da 4 a 6 volte nei 14 giorni), TC (Telefonata di Controllo).

I pazienti furono dunque randomizzati a ricevere uno dei 4 trattamenti:
a. MP + TT
b. MP + TC
c. PL + TT
d. PL + TC

Al giorno 1 (basale), al giorno 8 e al giorno 15, i pazienti compilarono i seguenti questionari:
- FACIT-F (Functional Assessment of Chronic-Illness and Therapy, composto da 27 domande sulla qualità di vita, articolate in 4 domini (fisico, sociale, emozionale e funzionale) + il modulo specifico per la fatigue (13 domande);
- ESAS (Edmonton Symptom Assessment Scale, una scala da 0 a 10), che valuta 9 sintomi comunemente manifestati da pazienti con tumore in fase metastatica;
- HADS (Hospital Anxiety and Depression Scale, 14 item) per la valutazione di ansia e depressione;
- PSQI (Pittsburg Sleep Quality Index) che valuta i disordini del sonno.

Endpoint primario: determinare se il miglioramento della fatigue, valutata con il FACIT-F, dal basale al 15° giorno, fosse superiore nei pazienti che ricevevano il metilfenidato rispetto a quelli trattati con placebo.
Poiché i dati non erano distribuiti normalmente, i confronti tra i 4 gruppi, valutando in ciascuno la differenza tra il basale e il giorno 15 (e tra il basale e il giorno 8), furono eseguiti comparando le mediane con il test di Wilcoxon per campioni indipendenti. Come endpoint secondari furono confrontate le differenze mediane nelle sottoscale del FACIT-F usando il test di Kruskall-Wallis. Inoltre, in ciascun braccio, fu analizzato l'andamento nel tempo (al giorno 8 e al giorno 15) di ciascuno dei parametri considerati rispetto al basale (giorno 1). Il test di Wilcoxon per dati appaiati fu usato per analizzare il miglioramento in HADS tra il giorno 1 e il giorno 15 separatamente nei pazienti trattati con MP e con PL.
Un campione di 212 pazienti avrebbe assicurato un'"adeguata potenza" per scoprire una differenza di almeno il 33% tra MP e PL nella valutazione della fatigue al giorno 15 rispetto al basale (endpoint primario).
Risultati. Dei 190 pazienti arruolati, 141 furono randomizzati ai 4 gruppi: 37 nel gruppo a), 31 in b), 38 in c) e 35 in d). Risultarono così valutabili 68 pazienti trattati con MP e 73 con PL; 75 nel gruppo TT e 66 in TC. Nessuna differenza significativa fu riscontrata al basale tra i 4 gruppi.
La mediana della fatigue, valutata sia con il FACIT-F che con l'ESAS, diminuì al 15° giorno rispetto al basale in tutti e 4 i gruppi, ma il miglioramento avvenne quasi del tutto entro l'8° giorno.
Considerando insieme TT e TC, la diminuzione riscontrata nel gruppo MP non risultò significativamente diversa da quella osservata nel gruppo PL, così come, mettendo insieme MP e PL, non risultarono significativamente diverse le differenze tra TT e TC.
Infine, il test di Kruskall-Wallis, per i 4 gruppi non risultò significativo.
Invece, la sintomatologia apprezzata con ESAS si rivelò mutare favorevolmente, in modo significativo, tra il basale e il giorno 8 e tra il basale e il giorno 15, per quasi tutti i sintomi (ad eccezione del dolore) e per la sensazione di benessere nei pazienti che ricevettero l'intervento telefonico (TT), ma fatigue, depressione e dispnea migliorarono significativamente anche nel corrispondente gruppo di controllo (TC).
Anche HADS e PSQI non mostrarono differenze significative tra il gruppo MP e il gruppo PL, così come non risultarono significativamente sbilanciati gli eventi avversi di grado ≥ 3. •

- Copyright - Il Pensiero Scientifico Editore downloaded by IP 3.149.243.32 Mon, 29 Apr 2024, 01:09:41

Negli anni '80 del secolo scorso, un importante statistico medico americano rivalutò i lavori pubblicati tra il 1960 e il 1979 sulle quattro più importanti riviste di medicina in relazione alla potenza degli studi pubblicati e trovò che gli studi negativi in circa i 2/3 dei casi non avevano una potenza sufficiente per evidenziare l'efficacia del nuovo trattamento. Concluse che in quel ventennio la medicina aveva perso molti farmaci potenzialmente utili. Non vorremmo che questa sia la ragione della negatività dei risultati ottenuti per il metilfenidato.

Un altro elemento negativo dello studio è che oltre ¼ dei pazienti arruolati non sono stati valutati (49/190). Quando la percentuale dei pazienti non valutati è alta, c'è sempre da temere la presenza di una distorsione da selezione (ad esempio, che siano stati valutati i pazienti in mediamente migliori condizioni e, pertanto a questi, e non a tutti i pazienti, siano riferibili i risultati dello studio). Inoltre, nessun cenno è stato fatto in relazione al criterio di valutazione considerato (se per intenzione a trattare, per protocol, per efficacia clinica).

Interessante è invece l'osservazione che i dati ottenuti con scale psicometriche, come quelle adottate nel lavoro, non possono essere ritenuti provenienti da variabili distribuite normalmente (cioè secondo la curva di Gauss). Questa considerazione, se da un lato ha indotto gli autori a condurre tutta l'analisi dei risultati con test non parametrici, dall'altro ha spinto noi ad approfondire l'argomento sulla scelta del test statistico da usare in relazione al tipo di dati e, quindi, alle scale di misura con cui sono stati ottenuti (v. in questo numero Statistica per concetti 2), rinviando l'argomento dei test non parametrici al prossimo numero di CASCO per motivi di spazio. Ma, come al solito in questo lavoro, le luci sono presto oscurate da ombre. Nella scelta dei test non parametrici come strumento di analisi dei risultati vi sono almeno due osservazioni da fare.

La prima è che sono stati eseguiti parecchi test statistici, senza che gli autori si siano preoccupati di apportare alcuna correzione ai livelli di significatività, ma esponendo il problema solo in sede di discussione, quando trattano dei limiti del lavoro: "Additional limitations include the fact that there was no statistical control for multiple comparisons, so some significant findings might be a result of type 1 error". Per loro fortuna i risultati relativi alla fatigue sono tutti negativi, per cui, su di loro, la correzione del livello di significatività non avrebbe prodotto alcun effetto.

A nostro avviso, in sede di discussione, la dichiarazione dei limiti (e spesso non tutti) di uno studio clinico, se da un lato attesta l'onestà intellettuale degli autori, dall'altro non minimizza affatto i difetti dello studio che continuano a permanere nonostante essi siano stati dichiarati. In alcuni casi, sarebbe vantaggioso leggere all'inizio l'elenco dei difetti per non perder tempo a proseguire nella lettura.

La seconda osservazione concerne il fatto che i dati ottenuti con un disegno fattoriale completo vanno analizzati con modelli lineari e, non potendolo fare (come in questo caso), occorre ricorrere a modelli lineari generalizzati (Generalized Linear Models, GLM) con cui è possibile trattare anche dati ordinali. Infatti, usando altri strumenti statistici (come i test non parametrici), si va inevitabilmente incontro alla correzione di Bonferroni (v. Statistica per concetti, CASCO, vol.1, n. 2, 2011: 24-25), per cui, tenendone conto, diventa assai difficile trovare differenze significative. L'uso di un test parametrico dovrebbe essere pertanto limitato ad ottenere la risposta al quesito principale dello studio, o poco più.

Il lavoro è senz'altro stimolante per via delle luci che presenta, ma come valutare i risultati ottenuti?

Per i numerosi difetti e punti non precisati, complessivamente, riteniamo che i risultati sull'inefficacia del metilfenidato non siano conclusivi, anche perché il periodo di osservazione (due settimane) potrebbe essere troppo breve per individuare un effetto non rapido (e nemmeno eclatante, dato che non sono state trovate differenze significative). Molti pazienti hanno avuto un miglioramento della fatigue indipendentemente dal trattamento, che però hanno comunque ricevuto, anche se si trattava di un placebo. Quindi, sulla base dei risultati negativi dello studio, non possiamo giungere alla conclusione (come del resto non fanno gli autori) che non occorra trattare il paziente, ma, al contrario, che sia necessaria ancora molta ricerca per giungere ad un soddisfacente controllo della fatigue.

Infine, per quanto concerne la telefonata terapeutica, si osservi che il suo effetto è stato confrontato con quello di una telefonata non terapeutica (di controllo), per cui tutti i pazienti hanno comunque ricevuto dalle 4 alle 6 telefonate nel periodo di trattamento. Forse lo strumento psicologico, in questa sede, non si è dimostrato di grande valore, ma, nella nostra esperienza, la telefonata in sé può rappresentare un prezioso strumento per rassicurare il paziente, per farlo sentire seguito dalla struttura che lo ha in cura. Il contatto continuativo della struttura con il paziente riteniamo sia comunque importante non solo per alleviare una sintomatologia, come la fatigue, ma anche, in altri campi, per ottenere risultati in termini di miglioramento del decorso della malattia (quanto meno migliorando la compliance del paziente alla terapia), che si tradurrebbe anche in risparmi sulla spesa sanitaria (ad es., evitando successivi ricoveri). In tutti i campi della medicina, l'effetto dell'intensificazione del management del paziente, ottenuta anche aumentando i contatti tra struttura e paziente, andrebbe indagato mediante studi di outcome research. •