

Agopuntura e fatigue

Enzo Ballatori
Statistico medico,
Spinetoli

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

Riassunto

L'articolo Molassiotis A, Bardy J, Finnegan-John J, et al. *Acupuncture for cancer-related fatigue in patients with breast cancer: a pragmatic randomized controlled trial*. *J Clin Oncol*, 2012; 30: 4470-6 è stato sintetizzato nella scheda e discusso in questa rubrica, focalizzando i difetti in esso contenuti, in particolare, la randomizzazione 3:1 e la possibile presenza di bias da selezione, la mancata adozione di tecniche di cecità, l'insufficienza del periodo di osservazione, la rilevanza clinica dei risultati ottenuti, l'analisi statistica condotta.

Parole chiave. Agopuntura, fatigue, randomizzazione, distorsione da selezione, distorsione da informazione, analisi della covarianza.

Summary

Acupuncture and fatigue

The paper Molassiotis A, Bardy J, Finnegan-John J, et al. *Acupuncture for cancer-related fatigue in patients with breast cancer: a pragmatic randomized controlled trial*. *J Clin Oncol*, 2012; 30: 4470-76 was first summarized (see "Scheda"), and subsequently discussed focusing on its shortcomings, like randomization 3:1 and possible selection bias, the lack of blindness, the insufficient period of observation, the clinical relevance of the obtained results, and the performed statistical analysis.

Key words. Acupuncture, fatigue, randomization, selection bias, information bias, covariance analysis.

Molta ricerca clinica è stata condotta sulla fatigue: si calcola che in oltre 170 studi, incentrati su interventi terapeutici, la fatigue sia stato l'endpoint principale o uno degli endpoint secondari (Bower JE. Treating cancer-related fatigue: the search for intervention that target those most in need. J Clin Oncol 2012; 30: 4449-50, editoriale). Si tratta, però, di una ricerca clinica sparsa, cioè vertente su innumerevoli possibili rimedi, che vanno dal cortisone¹, al ginseng², all'agopuntura (v. scheda), generalmente testati contro placebo. Malgrado nei lavori pubblicati questi trattamenti siano risultati quasi sempre efficaci, ancora non esiste una terapia standard contro la fatigue, probabilmente sia perché è modesto il loro effetto, sia perché mancano studi comparativi e di combinazione.

Per valutare la fatigue esistono numerosi strumenti, validati in molti paesi, e proposti in forma di questionari autosomministrati, come il Multidimensional Fatigue Inventory (MFI), il Brief Fatigue Inventory (BFI), il Fatigue Severity Scale (FSS), il modulo per la valutazione della Fatigue del Functional Assessment of Cancer Therapy (FACT-F), che può essere usato indipendentemente dal questionario generale (FACT-G), il Fatigue Symptom Inventory (FSI), che valuta anche la durata del sintomo, e così via.

Vi è ampio accordo sul fatto che quando si valutano comparativamente due terapie tese a controllare un sintomo la cui intensità è valutabile solo soggettivamente, occorre ricorrere a tecniche di cecità perché la conoscenza del trattamento somministrato potrebbe alterare, anche di molto, la risposta fornita dal paziente.

Il lavoro sintetizzato nella scheda è complessivamente ben confezionato, ma presenta alcuni problemi rilevanti ed altri minori che impediscono di accettare senza riserve le conclusioni.

a. Mancata adozione di tecniche di cecità

Come attuare la cecità quando il trattamento è l'agopuntura?

La risposta viene da lontano, da uno dei primi trial volti a provare l'esistenza dell'effetto placebo (anni '60, ci scusiamo con il lettore per l'imprecisione del ricordo). Alcune decine di pazienti tutti sofferenti di mal di schiena furono randomizzati a ricevere l'agopuntura, una pressione dell'ago (ma senza infiggerlo), nulla. I risultati furono che i primi due trattamenti ottennero una riduzione del dolore superiore a quella riscontrata nel gruppo di controllo, ma paragonabile tra loro. Si potrà obiettare sulla scarsa potenza dello studio, ma certo l'effetto dell'agopuntura, in quel contesto, non fu comunque esaltante.

Non è chiaro il motivo per cui, nello studio considerato, ai pazienti del gruppo di controllo non venne praticata una finta agopuntura; certo, non per ragioni etiche, in quanto, nel consenso informato, si sarebbe potuto scrivere che i $\frac{3}{4}$ dei pazienti sarebbero stati sottoposti ad agopuntura, mentre ad $\frac{1}{4}$ di loro sarebbe stato infisso l'ago ad una profondità non utile per risultare efficace. Infatti, non solo sapere di ricevere un trattamento potenzialmente attivo potrebbe indurre di per sé una risposta positiva, ma lo stesso contatto con l'operatore avrebbe potuto migliorare la fatigue del paziente.

b. Randomizzazione 3:1 e distorsione da selezione

Non abbiamo mai visto una randomizzazione 3:1 al di

SCHEDA

Molassiotis A, Bardy J, Finnegan-John J, et al. Acupuncture for cancer-related fatigue in patients with breast cancer: a pragmatic randomized controlled trial. J Clin Oncol 2012; 30: 4470-6.

Si tratta di una sperimentazione clinica controllata condotta su donne affette da carcinoma mammario, con fatigue persistente e almeno moderata, in cui le partecipanti furono randomizzate nel rapporto 3:1 a ricevere, rispettivamente, 6 sedute settimanalmente di agopuntura o un libretto illustrativo sul management della fatigue.

Per selezionare le pazienti con i suddetti livelli di fatigue fu considerato un valore non inferiore a 5 di un analogo visivo a 10 punti, messo a punto dallo stesso gruppo di studio, che si correla con una scala psicometrica di misura della fatigue ($r = 0,75$).

Furono considerate eleggibili anche le pazienti con recidiva locale, ma vennero escluse quelle con metastasi a distanza.

Intervento. Tutte le pazienti ricevettero le cure usuali per la fatigue. Mancando uno standard, queste vennero decise dai medici che le seguivano.

La pazienti del gruppo sperimentale furono sottoposte a 6 sedute settimanali di agopuntura della durata di 20' ciascuna raccomandando agli agopunturisti di agire su 3 punti precisati, ma non vincolandoli ad essi. Gli aghi furono infissi ad una profondità variabile da 1 a 2,5 cm, ed ogni conversazione tra paziente ed agopunturista fu tenuta al minimo. Furono coinvolti 12 agopunturisti con almeno 2 anni di esperienza clinica.

Il rapporto di randomizzazione 3:1 fu scelto per consentire altri studi sull'utilità di una successiva agopuntura o di una ago-pressione

autosomministrata dalla paziente. Le valutazioni furono eseguite dopo 6, 10 e 18 settimane, ma quelle relative a 10 e 18 settimane sono riportate in un altro lavoro.

Outcome. La differenza del punteggio complessivo del Multidimensional Fatigue Inventory (MFI) dopo 6 settimane, rispetto al basale, fu considerata come endpoint primario.

L'MFI è un questionario psicometrico autosomministrato, composto da 20 item, in grado di valutare la fatigue in generale (punteggio complessivo), ma anche le singole dimensioni della fatigue fisica, mentale, l'attività e la motivazione.

Come endpoint secondari furono valutate, oltre le sottoscale dell'MFI, anche l'ansia e la depressione, usando il questionario Hospital Anxiety and Depression Scale (HADS, con 14 item), nonché la qualità di vita misurata con il Functional Assessment of Cancer Therapy-General (FACT-G) ed il modulo per il carcinoma mammario (FACT-B).

Analisi statistica. L'endpoint primario fu analizzato con modelli di analisi di covarianza, assumendo il punteggio dell'MFI alla sesta settimana come variabile dipendente, il punteggio basale come covariata e il trattamento come fattore. Anche per gli endpoint secondari fu usata un'analisi simile con strategie di Last Observation Carried Forward (LOCF) per dare una valutazione ai missing values così da assoggettare i risultati ad un'analisi di sensibilità. Furono anche usati t-test per confronti tra medie di entrambi i gruppi. Il costante ricorso a procedure parametriche fu giustificato con il teorema centrale del limite, per cui, indipendentemente dalla distribuzione di partenza, nel caso di grandi campioni, le distribuzioni

campionarie degli stimatori tendono alla normale.

La dimensione del campione, descritta in un precedente lavoro, fu fissata in 320 pazienti, 240 nel gruppo di trattamento ed 80 in quello di controllo.

Risultati. Furono randomizzati 302 pazienti, 227 al trattamento sperimentale e 75 alle usuali cure potenziate dalla distribuzione di un libretto sulla fatigue.

La valutazione finale dell'endpoint primario non fu rinviata al centro sperimentatore da 10 pazienti del gruppo di controllo (13,3%) e da 46 pazienti del gruppo di trattamento (20,3%). In quest'ultimo, 9 pazienti non furono trattati e 53 ebbero un numero di sedute inferiore a 6. Questi ultimi furono comunque valutati con il criterio LOCF in base al principio di intenzione a trattare. In conclusione l'analisi fu eseguita su 181 pazienti nel gruppo sperimentale e 65 in quello di controllo.

Per l'endpoint primario, la differenza tra l'effetto (punteggio complessivo della fatigue alla sesta settimana meno lo stesso punteggio al basale) nel gruppo di agopuntura vs quello di controllo fu $-3,11$ (dati non aggiustati per il basale: $-3,72$ nel gruppo di agopuntura e $-0,62$ in quello di controllo), con un intervallo di confidenza al 95% che va da $-3,97$ a $-2,25$ ($P < 0,001$). I risultati furono accompagnati da una conservativa analisi di sensibilità condotta ipotizzando, per i pazienti che non completarono le 6 settimane previste di agopuntura, che nessun miglioramento dell'MFI si sarebbe registrato dall'ultima osservazione eseguita alla sesta settimana. I risultati, sebbene alquanto attenuati, restarono significativi.

Analogamente, risultati significativi a favore dell'agopuntura si ottennero anche per gli endpoint secondari. •

fuori di un contesto in cui il trattamento attivo era noto avere una straordinaria efficacia in confronto a quello standard, ma non sembra questo il caso, e comunque mai quando la risposta è di tipo soft (non la sopravvivenza, per intenderci). Il risultato è che il gruppo di controllo consta di soli 65 pazienti: è vero che c'è una randomizzazione, è vero che si sta valutando la differenza con il basale, ma quando si tratta con piccoli numeri tutto può accadere, anche per solo effetto del caso e, comunque, i risultati sono poi difficilmente generalizzabili. Gli autori giustificano la randomizzazione 3:1 con il fatto che i risultati presentati si riferiscono ad uno spezzone di una ricerca più articolata, per cui, complessivamente, questo era il rapporto di randomizzazione più indicato.

Il quadro peggiora avendo constatato che circa il 13% dei pazienti del gruppo di controllo ed il 20% di quelli trattati non sono stati valutati. Nel lavoro non c'è alcuna spiegazione, né v'è traccia di un tentativo volto alla loro identificazione in relazione a determinate caratteristiche connesse alla fatigue (come, ad es., la durata pregressa del sintomo, l'età, lo stadio della malattia, le loro attività abituali prima di ammalarsi, il supporto familiare, e così via).

c. Che succede dopo?

La fatigue è un sintomo di norma assai persistente, per cui valutare l'efficacia dell'agopuntura dopo sole 6 settimane di trattamento, senza curarsi di ciò che accade dopo, apre la porta a numerosi possibili scenari di cui non sappiamo nulla (come ad es., un peggioramento della fatigue per i soggetti che hanno ricevuto l'agopuntura, oppure un loro sostanziale miglioramento).

Si noti che l'agopuntura è stata praticata una volta a settimana (20' a seduta), per 6 settimane, senza un serio studio preliminare sull'ottimizzazione del timing e della durata del trattamento. In altre parole, non è noto se facendo più sedute settimanali migliora la risposta osservata o meno, né se praticando l'agopuntura per più di 6 settimane, il sintomo migliora ulteriormente, o no.

d. Risultato significativo, ma clinicamente rilevante?

Lo strumento utilizzato, l'MFI, consta di 20 item. La risposta al trattamento è valutata come differenza tra il punteggio complessivo dell'MFI dopo 6 settimane e il basale. La differenza media tra i due bracci è stata pari a 3,10 a favore dell'agopuntura, ma non è dato di sapere quale sia l'interpretazione clinica di tale decremento, cioè quanto sia stato rilevante per il paziente.

e. Studio pragmatico

Quando s'incontra l'aggettivo "pragmatico" associato ad uno studio clinico, si percepisce un pericolo dovuto al rumore introdotto in eccesso da qualche variabile che non si è voluto, o non si è riusciti, a controllare. La mente rimanda ai celebri studi pragmatici "GISSI" in cui decine di migliaia di pazienti furono arruolati non solo per percepire una piccola differenza di efficacia, ma anche per tener conto del rumore introdotto dalla pratica clinica lasciata volutamente libera (ad eccezione del trattamento in studio). Quindi, per fare in

modo che il rumore non sovrasti il segnale, uno studio pragmatico in genere si basa su un ampio numero di pazienti arruolati, mentre, nel nostro caso, sono meno di 250.

Ma cos'ha di pragmatico lo studio? Dov'è l'eccesso di rumore non controllato?

Anzitutto il trattamento sperimentale, eseguito da 12 agopunturisti diversi, che non è stato standardizzato (e forse non poteva esserlo). Tre punti di infissione sono stati indicati, ma poi l'agopunturista è stato lasciato libero di operare come meglio credeva nell'interesse del paziente. Quindi, i pazienti del gruppo sperimentale hanno ricevuto trattamenti diversi, dando luogo ad una probabilmente debole riproducibilità (non necessariamente in negativo), nel caso lo studio fosse replicato con altri agopunturisti.

In secondo luogo, le altre terapie contro la fatigue, consistenti sia in farmaci, sia in raccomandazioni in termini di mutamento di stili di vita, sono state lasciate libere, per cui una maggiore insistenza su di esse nei pazienti che settimanalmente erano sottoposti ad agopuntura potrebbe aver contribuito alla positività del risultato. Si badi bene, non è detto che una maggiore insistenza vi sia stata, ma che poteva esserci semplicemente perché nel braccio sperimentale il paziente è venuto più spesso a contatto con la struttura sanitaria.

Vi sono anche obiezioni minori che possono essere sollevate allo studio in esame.

Una di queste riguarda l'uso dell'analisi di covarianza che, com'è noto, poggia su due ipotesi, una delle quali è la normalità della distribuzione dell'errore. A prima vista, è abbastanza inverosimile che una valutazione condotta con scale di Likert a 10 punti possa essere ritenuta normale (il punteggio va da 1 a 10 ed è discreto, cioè non varia con continuità, mentre la curva normale è continua ed asintotica; in un questionario psicometrico, inoltre, la distanza tra 1 e 2, ad esempio, non è uguale a quella che c'è tra 3 e 4, o tra 9 e 10). Per giustificare l'ipotesi di normalità, gli autori fanno appello al teorema centrale del limite, per cui, semplificando, qualunque sia la distribuzione di partenza, al tendere di n (dimensione del campione) all'infinito, essa converge ad una normale.

Purtroppo, però, si tratta solo di un teorema limite, che se assicura la convergenza alla normale quando n tende all'infinito, non garantisce la validità per campioni di qualche centinaio di unità. In negativo, se la pretesa degli autori di poter trattare come normali variabili che non lo sono mediante il ricorso al teorema centrale del limite (vecchio di quasi un secolo: Lindeberg, 1922) fosse stata considerata accettabile, non ci sarebbero state le successive innovazioni nella Statistica, come i test non parametrici per scale ordinali (l'MFI è una di queste), come il tentativo di normalizzare distribuzioni che non lo sono attraverso opportune trasformazioni matematiche dei dati (ad es., la trasformazione radice, o quella logaritmica), come l'introduzione di modelli lineari generalizzati per dati che, appunto, hanno una distribuzione differente da quella normale. Il ragionamento degli autori è comunque vecchio, in quanto, oggi, una tale assunzione andrebbe provata a posteriori mediante l'analisi dei residui, che, invece, non è stata condotta.

Ci si chiede infine perché non sia stato usato uno stru-

mento più semplice e a portata di mano (nel senso che qualunque software statistico è in grado di calcolarlo), come un test non parametrico (ad es., il test U di Mann-Whitney) per il confronto tra due mediane nel caso di dati ordinali. Il presupposto c'era: la randomizzazione, che assicura un buon bilanciamento dei dati ivi incluso il punteggio basale dell'MFI, rende inutile aggiustare per tale variabile. In conclusione, sarebbe stato sufficiente eseguire un test U sulla risposta alla sesta settimana per valutare l'efficacia differenziale dei due trattamenti. Certo, l'effetto sul lettore sarebbe stato forse meno impressionante.

Conclusioni

È difficile usare metodi scientifici per valutare l'efficacia delle medicine alternative che, a differenza della medicina occidentale, hanno un approccio olistico che coinvolge categorie trascendenti, facendo riferimento non solo all'atto materiale del trattamento, ma invocando anche proiezioni in un universo astrale, considerato un tutt'uno con il paziente. È difficile, ma occorre tentare per individuare possibili benefici in esse contenuti. Sotto questo profilo, il lavoro esaminato è lodevole, sebbene inficiato da diversi problemi che forse potevano essere evitati con una migliore pianificazione dello studio.

Randomizzazione 3:1. La ricerca clinica è un'attività complessa. A nostro avviso, è sempre opportuno cercare di evitare di aggiungere ulteriore complessità raggruppando studi aventi diverse finalità in un'unica ricerca. In altre parole, è sempre auspicabile porsi una pluralità di domande, ma non necessariamente le risposte vanno ottenute con un solo studio.

Distorsione da selezione. Nel braccio di controllo il 13% dei pazienti non ha restituito l'MFI compilato, nel braccio di trattamento il 20%. La differenza non è significativa, ma sarebbe stato opportuno approfondire le ragioni della mancata risposta. Se consideriamo che la dimensione del campione è già esigua (soprattutto nel braccio di controllo), che numerosi pazienti del gruppo sperimentale non hanno avuto il trat-

tamento (correttamente sono stati ugualmente valutati in accordo al principio di intenzione a trattare) o lo hanno avuto incompleto, e che il trattamento non è stato lo stesso per tutti i pazienti trattati, a nostro avviso l'approfondimento sulle cause dei drop out sarebbe stato più che opportuno.

Oltre che alla sesta settimana, i pazienti sono stati valutati anche alla decima e alla diciottesima settimana: è veramente increscioso non avere informazioni su come siano andate successivamente le cose. La fatigue è un sintomo persistente, per cui sarebbe opportuno conoscere la sua dinamica anche oltre il periodo considerato (18 settimane, cioè quattro mesi e mezzo): speriamo di apprendere dai prossimi lavori che si sia proseguito nel monitoraggio dei pazienti anche una volta terminato lo studio.

A nostro avviso, la mancata adozione di tecniche di cecità nella somministrazione del trattamento del gruppo sperimentale è un importante elemento che rende poco affidabili le conclusioni dello studio, anche perché va ad aggiungersi agli altri problemi sopra esaminati, in particolare a quelli che rendono "pragmatico" lo studio (diversi agopunturisti con una differente preparazione, mancata standardizzazione dell'agopuntura, mancato controllo delle altre terapie somministrate contro la fatigue).

Auspichiamo che vengano condotti ulteriori studi sull'effetto dell'agopuntura come terapia per la fatigue, magari anche per altri tipi di neoplasie, possibilmente in confronto e/o in aggiunta ad altri trattamenti già trovati efficaci nell'attenuare questo sintomo così distruttivo. •

Bibliografia

1. Yennurajalingam S, Frisbee-Hume S, Palmer JL, et al. *Reduction of cancer-related fatigue with dexamethasone: a double-blind randomized, placebo-controlled trial in patients with advanced cancer.* *J Clin Oncol* 2013; 31: 3076-82.
2. Barton DL, Liu H, Dakhil SR, et al. *Wisconsin ginseng (Panax quinquefolius) to improve cancer-related fatigue: a randomized, double blind trial, No7C2.* *J Natl Cancer Inst* 2013; 105: 1230-7.