Pillole di EBM - Capitolo 5

Data 01 settembre 2006
Categoria scienze_varie

La numerosità del campione e la durata degli studi.

La numerosità del campione

Nei capitoli precedenti ci siamo occupati di alcuni aspetti che bisogna considerare quando si valuta uno studio clinico. Il numero dei pazienti arruolati è un altro parametro importante e facile da reperire già dall'abstract. E' intuitivo che tanto più numeroso è il campione arruolato più i risultati dovrebbero essere validi e affidabili.
La numerosità del campione è importante perché spesso solo in tal modo si può avere la potenza statistica per rilevare determinati eventi. E' evidente che se un trattamento ha lo scopo di ridurre un evento che già di per sé non è frequente bisogna arruolare molti pazienti (migliaia o decine di migliaia) per poterlo rilevare. La numerosità del campione viene quindi ritenuta, di solito, sinonimo di studio clinico importante. Questo però può portare anche a delle incongruenze. Supponiamo per esempio che un farmaco riduca l'ictus rispetto al farmaco concorrente e che però la differenza possa diventare statisticamente significativa (vedremo in seguito cosa significa questa espressione) solo se si studia un numero molto elevato di pazienti (per esempio 30-40.000). Succede allora che differenze marginali vengono amplificate perché si sono reclutati moltissimi soggetti: si ottiene una significatività statistica ma l'utilità clinica è discutibile. In effetti si assiste sempre più spesso alla organizzazione di mega-trial che hanno lo scopo di mettere in evidenza differenze di efficacia tra due farmaci molto piccole, che non risulterebbero se la casistica fosse meno numerosa. Aumentare la casistica può quindi essere un escamotage per evidenziare benefici piccoli.
Mi direte: ma scusa se un farmaco è efficace lo è tanto sui piccoli quanto sui grandi numeri. La faccenda purtroppo non funziona così. Per farmi capire farò un esempio perché penso che gli esempi servano molto di più di tante parole. Dovrò purtroppo usare dei numeri (anche se cercherò di usarli molto semplici) e alcuni concetti che per ora a molti possono non essere chiari. Lo diventeranno in seguito, per il momento l'importante è seguire il filo del ragionamento.

Supponiamo di avere un nuovo farmaco che riduce il rischio di infarto e di volerlo confrontare con un farmaco già ampiamente usato. Arruolo quindi 5.000 soggetti e li divido in maniera randomizzata in due gruppi di 2.500 ciascuno.
Al primo gruppo somministro il nuovo farmaco, al secondo gruppo quello più vecchio. Dopo cinque anni vado a vedere quanti infarti ci sono stati nei due gruppi.
Ecco i risultati:

·Gruppo farmaco nuovo: 2.500 trattati e 140 infarti
·Gruppo farmaco vecchio: 2.500 trattati e 160 infarti

Sembra quindi che il nuovo farmaco sia più efficace del vecchio, ma la statistica è una cosa strana, non basta dimostrare che si hanno meno infarti, bisogna anche dimostrare che tale riduzione non è dovuta semplicemente al caso ma, come si dice, è **significativa dal punto di vista statitistico** . Per fare questo si calcola dapprima la riduzione relativa del rischio di infarto avuta con il nuovo farmaco e si trova che questa è del 12,5% (in effetti se si fa 160 meno 12,5% si avrà 140). Più avanti impareremo anche a calcolare questa riduzione (non è così difficile come potrebbe sembrare). Il che vuol dire che se facciamo per comodità uguale ad 1 il rischio di infarto con il vecchio farmaco, il rischio col nuovo farmaco sarà di 1 meno 12,5% = 0,875. Impareremo più avanti che questo numero viene chiamato **rischio relativo** e si esprime con la sigla RR = 0,875. Però il calcolo del rischio relativo non basta ancora, bisogna anche trovare il cosiddetto **intervallo di confidenza al 95%** (che si scrive così: IC95%). Infatti un risultato per essere significativo dal punto di vista statistico deve avere un IC95% che **non comprende l'unità** , cioè il numero 1. Se per sfortuna lo comprende vuol dire che la differenza trovata conta poco o nulla perché NON E' SIGNIFICATIVA STATISTICAMENTE. Nel caso dell'esempio in esame l'IC95% del RISCHIO RELATIVO va da 0,7 a 1,09 e comprende perciò il numero 1. Ora non importa sapere come si è arrivati a trovare tale intervallo, quello che importa è che il 12,5% in meno di infarti trovati con il nuovo farmaco non conta nulla e che i due trattamenti devono essere considerati di efficacia paragonabile. Non state neppure a chiedervi per ora come mai l'IC95% non deve comprendere il numero 1 e credetemi sulla parola. Quando tratteremo l'argomento potete tornare a rileggervi questo punto e tutto vi sembrerà liscio comel'olio.
Anzi per complicare un po' le cose riporto la tabella 1 che probabilmente per molti sarà poco comprensibile ma che mostra come in generale vengono riportati i risultati di uno studio.Intanto cominciate ad abituarvi a questo nuovo linguaggio, anche se non è molto chiaro il significato di alcuni termini (rischio relativo, intervallo di confidenza)

Tabella1

FARMACO NUOVO: 2500 TRATTATI, 140 INFARTI
FARMACO VECCHIO: 2500 TRATTATI, 160 INFARTI
RISCHIO RELATIVO (RR) = 0,875; INTERVALLO DI CONFIDENZA AL 95% (IC95%) = 0,70-1,09

Ovviamente un risultato del genere può non far comodo a chi produce il nuovo farmaco, allora che fa?
Arruola ben 30.000 pazienti, dividendoli poi in due bracci di 15.000 ciascuno. Somministra come di prassi nuovo e vecchio farmaco e dopo 5 anni tira le somme. Ecco i risultati:

·Gruppo nuovo farmaco: 15.000 trattati e 840 infarti
·Gruppo vecchio farmaco: 15.000 trattati e 960 infarti

Se anche qui andiamo a calcolare di quanto è stata la riduzione del rischio di infarto con il nuovo farmaco troveremo che essa è sempre del 12,5% (infatti 960 meno 12,5% fa appunto 840). Cioè l'RR (RISCHIO RELATIVO) sarà sempre 0,875. Ma andiamo adesso a vedere com'è l'intervallo di confidenza al 95%: troveremo che esso va da 0,80 a 0,96 e in questo caso non comprende il numero 1. **Pertanto la differenza trovata tra nuovo e vecchio farmaco è diventata miracolosamente importante e significativa** .
Nella tabella 2 vengono riportati in sintesi i risultati.

Tabella2

FARMACO NUOVO: 15.000 TRATTATI, 840 INFARTI
FARMACO VECCHIO: 15.000 TRATTATI, 960 INFARTI
RR = 0,875; IC95% 0,80-0,96

Tutto questo ha degli evidenti vantaggi nel senso che permette di scoprire differenze che altrimenti, con meno casi arruolati, non si potrebbero vedere, ma porta d'altra parte ad enfatizzare benefici che clinicamente potrebbero essere scarsi. L'organizzazione di mega-trial con decine di migliaia di pazienti arruolati ha lo scopo appunto di evidenziare differenze nei trattamenti che altrimenti non potrebbero risultare. Queste differenza acquistano importanza in termini statistici di popolazione quando vengono trattati milioni di persone, ma possono essere minime per il singolo paziente.
Nel caso esemplificato, se volessimo tradurre il tutto in termini facilmente comprensibili, potremmo anche dire che bisogna trattare 125 soggetti per 5 anni con il nuovo farmaco per avere un infarto in meno. Ciò vuol dire anche che per 5 anni ne tratto inutilmente 124, i quali assumeranno il farmaco per non avere nessun beneficio in più e correranno il rischio di eventuali effetti collaterali. Questo modo di vedere le cose fa riferimento al cosiddetto **NNT (numero di soggetti che è necessario trattare per evitare un evento in un determinato lasso di tempo)** . Impareremo in seguito anche a calcolarlo, questo famoso NNT, per ora mi serve solo per dire che il nuovo farmaco potrebbe essere utile solo se non porta ad un maggior numero di effetti collaterali gravi rispetto al vecchio farmaco. Se per ipotesi avessi ogni 125 pazienti trattati per 5 anni un infarto in meno rispetto al vecchio trattamento ma 3 uremie terminali in più, il beneficio andrebbe a farsi benedire. In altre parole il rapporto rischi/benefici non sarebbe favorevole.
Questo è un punto molto importante: se il disegno del trial non prevedesse di registare anche le uremie terminali potrei non aver ben chiaro il reale profilo di sicurezza del farmaco e ritenere che sia efficace perchè riduce gli infarti, ma non pericoloso per il rene. Purtroppo non sempre i trials sono disegnati per registare TUTTI gli eventi avversi gravi (SAEs = Serious Adverse Events)che si verificano, anche quelli che apparentemente non sembrano legati al trattamento testato, e questo porta a non avere un quadro chiaro del reale impatto sulla salute del farmaco in prova.
Un altro aspetto da considerare poi è il costo della nuova terapia: **se questo fosse molto elevato vale la pena investire molti soldi per avere un infarto in meno ogni 125 trattati per 5 anni o è preferibile investire i fondi in altri progetti sanitari con un costo ed un'efficacia più favorevoli?** . Come si può intuire le risposte non sono affatto semplici ma quando si giudica dell'utilità di un farmaco vi sono molti aspetti da considerare, non ultimi quelli economici, soprattutto in tempi di risorse sanitarie limitate.

La durata dello studio

La durata dello studio (detta anche follow-up) dipende naturalmente dall'end-point che si vuol misurare. Per esempio se si vuol vedere se un farmaco antipertensivo è in grado di ridurre la pressione più del placebo può bastare un follow-up di qualche mese. Se al contrario si vuol valutare se lo stesso farmaco è in grado di ridurre le complicanze della malattia ipertensiva (come per esempio l'ictus o l'infarto o lo scompenso cardiaco) è necessario disegnare uno studio con un follow-up adeguato, della durata di almeno qualche anno. Quando si valuta uno studio bisogna quindi sempre chiedersi se la durata dello studio è adatta a valutare gli outcomes previsti.
Per esempio in uno studio si vuol valutare l'efficacia di un farmaco proposto per il morbo di Alzheimer. Come end-point si sceglie la valutazione dello stato funzionale fisico e psichico misurata con un questionario prima dell'inizio dello studio e dopo 6 mesi. E' evidente che per una malattia a decorso cronico e progressivo come l'Alzheimer la valutazione dell'efficacia della terapia a sei mesi è probabilmente insufficiente a determinare se il farmaco è o meno efficace nel ridurre la disabilità e la progressione del morbo.
Come altro esempio porterò uno studio che aveva randomizzato meno di 200 donne a terapia ormonale sostitutiva o placebo. Dopo 12 mesi nel gruppo in terapia attiva si osservò una riduzione della proteina C reattiva (PCR) rispetto al gruppo di controllo. Siccome la PCR elevata è ritenuto un fattore di rischio cardiovascolare, si può ipotizzare che gli ormoni siano utili nella prevenzione cardiovascolare. A questo punto allora una domanda facile per i lettori: secondo voi è corretto, sulla base di questo studio, concludere che la terapia ormonale sostitutiva (TOS) protegge il cuore? Credo che chiunque mi abbia seguito, anche distrattamente, fino a questo punto non troverà difficoltà a rispondere che non è assennato trarre conclusioni di questo tipo. In effetti il supposto ruolo protettivo della TOS sulle malattie cardiovascolari è stato smentito clamorosamente dallo studio WHI che aveva arruolato più di 16.000 donne con un follow-up di circa 5

anni . Il confronto tra lo studio precedente e il WHI non è neppure pensabile, non solo perché il primo ha valutato un end-point surrogato e il secondo degli end-point clinici ben più importanti (mortalità, infarto, ictus, tromboembolismo venoso, ecc), ma anche per la numerosità del campione e la diversa durata.

RenatoRossi