Pillole di EBM - Capitolo 4

Data 01 agosto 2006
Categoria scienze_varie

Quali sono gli end-point di uno studio?

Siamo arrivati finalmente a parlare degli end-point (detti talora anche **esiti** o **outcomes**).
Sicuramente li avrete sentiti nominare molte volte e magari vi sarete anche chiesti che cosa diavolo sono.
Ebbene gli end-point non sono altro che "QUELLO" che alla fine dello studio i ricercatori si sono proposti di valutare. Con alcuni esempi mi spiego subito. Supponete di essere dei ricercatori che vogliono valutare l'efficacia di un farmaco antipertensivo. Allora misurerete la pressione ai vostri pazienti prima della somministrazione del farmaco e dopo. La riduzione media della pressione che otterrete potrebbe essere l'end-point dello studio.
Facciamo un altro esempio. State provando un farmaco per l'osteoporosi, confrontandolo con il placebo. Dopo cinque anni andate a contare quante fratture ci sono state. Quello è l'end-point dello studio.
Come si può capire da questi due esempi banali, sostanzialmente esistono due tipi di end-point: quelli **"hard"** (per esempio numero di decessi, infarti, ictus, ricoveri per scompenso cardiaco, fratture di femore, eccetera) e quelli cosiddetti **"surrogati"** (per esempio la pressione arteriosa, il colesterolo, la massa ossea, ecc).
E' intuitivo che gli end-point "hard" sono quelli maggiormente utili nel determinare il beneficio clinico di un trattamento. Ma allora, mi direte, perché in molti studi si trovano end-point surrogati?
La ragione sta nel fatto che questi ultimi sono più facili da ottenere e soprattutto richiedono un follow-up più breve: se sto provando un farmaco ipocolesterolemizzante, un conto è andare a vedere dopo 3 mesi quale è stata la riduzione del colesterolo, un conto è aspettare 5 anni per vedere se sono diminuiti gli infarti.
Gli end-point surrogati sono quindi molto comodi per i ricercatori, li si usa perché si ritiene che in qualche modo siano correlati agli outcomes "hard": è ragionevole infatti pensare che se riduco il colesterolo riduco anche il rischio di infarto.
Però purtroppo la medicina è una scienza un po' particolare, dove logica e ragionevolezza non sempre la fanno da padroni. In altre parole non è detto che ad un miglioramento di un end-point surrogato corrisponda un beneficio clinico. Vale quindi la regola che per giudicare realmente la bontà di un farmaco gli end-point surrogati non possono sostituire quelli hard.
Per capirci faremo due esempi.
Il primo riguarda uno studio che ha valutato l'efficacia della terapia ormonale sostitutiva nel migliorare l'assetto lipidico (lo studio è noto con la sigla PEPI). In questo lavoro circa 900 donne furono suddivise in 4 gruppi, in tre gruppi si usavano varie combinazioni di ormoni e nel quarto gruppo il placebo. Al termine dello studio venne evidenziato che la terapia ormonale sostitutiva riduceva i valori di colesterolo LDL e aumentava quelli dell'HDL. Bene, direte voi, ecco qui un trattamento che oltre a migliorare i sintomi della menopausa riduce il colesterolo cattivo e fa aumentare quello buono, abbiamo quasi trovato la "pallottola magica" per prevenire la cardiopatia ischemica nella donna proprio in un' età in cui essa diventa più vulnerabile a questa patologia. Apparentemente sembra tutto logico e ragionevole, ma purtroppo non è così. Quando poi vennero effettuati studi sulla terapia ormonale sostitutiva con end-point clinici hard (studio WHI, studio HERS, studio ESPRIT), sia in donne sane che cardiopatiche, si è visto che la terapia ormonale proposta per la menopausa non solo non protegge dalla cardiopatia ischemica, ma addirittura ne aumenta il rischio.
Il secondo esempio è forse ancora più eclatante e suggestivo e viene citato in tutti i testi di metodologica degli studi clinici. Si tratta dello studio CAST, in cui venne sperimentata la flecainide (un antiaritmico) nei soggetti infartuati con aritmie ventricolari. E' noto che nei pazienti post-infartuati vi è un aumento del rischio di morte improvvisa, specialmente nei primi giorni e nelle prime settimane dopo l'evento acuto. I soggetti più a rischio sono quelli che mostrano aritmie ventricolari frequenti e minacciose (fenomeno R/T, run di tachicardia ventricolare, BEV polimorfi, ecc.). E' logico e ragionevole pensare che se riuscissi, con una terapia farmacologica, a sopprimere o ridurre di intensità e frequenza queste aritmie ne avrei un vantaggio anche clinico, nel senso che avrei meno decessi per morte aritmica improvvisa.
Lo studio CAST dimostrò che il farmaco antiaritmico agiva positivamente sulle aritmie provocandone una netta riduzione. Tuttavia il trial venne sospeso anticipatamente. Come mai? L'analisi ad interim dei dati mostrava infatti un eccesso di morti nel gruppo trattato attivamente rispetto al gruppo di controllo che assumeva placebo. Questo studio è veramente esemplare in quanto dimostra elegantemente che un farmaco, che ha un effetto positivo su un end-point surrogato (aritmie), non necessariamente produce un beneficio clinico. Qualcuno parlò allora di "effetto cosmetico" del farmacosull'elettrocardiogramma.

Ho appena accennato alla cosiddetta **analisi ad interim** . Di cosa si tratta? E' una procedura di garanzia messa in atto per evitare brutte sorprese. Mentre il trial è in corso i dati preliminari vengono costantemente monitorati da un comitato di esperti in modo da interrompere lo studio prima del termine previsto se il numero di eventi predefiniti (infarti, ictus, ricoveri per scompenso, fratture femorali, ecc.) dovesse superare una certa soglia rispetto all'altro braccio. Lo studio può essere sospeso anticipatamente sia perché il farmaco si dimostra "troppo" efficace rispetto al controllo, sia nel caso opposto, quando il farmaco si dimostra meno efficace. Nel primo caso non sarebbe etico continuare lo studio privando i malati (sia quelli arruolati nel braccio controllo sia tutti i malati in genere) di un trattamento che si è dimostrato chiaramente utile anche prima che lo studio finisca, nel secondo caso non è giusto continuare a trattare soggetti con un farmaco che fa peggio del controllo.
Tuttavia l'interruzione precoce di una studio comporta anche un rovescio della medaglia. Supponiamo per esempio che si sia reso disponibile un nuovo trattamento per i pazienti sieropositivi per HIV. Viene effettuato uno studio che paragona

la nuova terapia con quella già disponibile, per determinare se si riesce a ritardare la comparsa di AIDS conclamato. Lo studio dovrebbe durare 6 anni ma dopo 3 viene interrotto perché l'analisi ad interim mostra un numero molto minore di insorgenza di AIDS con il farmaco nuovo. Tuttavia ci si accorge che questo farmaco provoca anche un aumento dell'ictus e dell'infarto fatali, ma la differenza non è significativa rispetto al farmaco di confronto. Non si può escludere però che tale differenza avrebbe potuto diventare rilevante se lo studio fosse durato i sei anni pianificati. In questo caso l'interruzione anticipata da un lato permette di estendere i benefici del nuovo farmaco nel ritardare la comparsa di AIDS a una vasta schiera di malati, dall'altro impedisce di valutare compiutamente il suo profilo di rischio.

Un ulteriore aspetto da considerare a proposito degli end-point è quello dei **criteridiagnostici** .
Essi devono essere predefiniti in modo accurato. Che cosa voglio dire con questo?
Se lo studio, poniamo, si propone di valutare di quanto migliorerà l'artrite reumatoide con un farmaco biologico, si dovranno stabilire dei criteri oggettivi di attività della malattia (in genere determinati con un sistema a punteggio che dovrà essere usato da tutti i medici partecipanti alla ricerca) sia al baseline che al termine del follow-up.
E' evidente che per alcuni end-point non è necessario predefinire nulla. Per esempio se si valutano i decessi, questi sono decessi punto e basta, non si corre certo il rischio che un medico usi dei criteri diversi di diagnosi. Un caso particolare però è rappresentato dall'end-point che valuta non i decessi in sé ma i decessi specifici (per esempio decessi per scompenso, per cancro mammario, ecc.) perché in questa evenienza si potrebbe avere una errata classificazione, detta misclassificazione dell'end-point. Per esempio in uno studio che volesse determinare la mortalità da cancro prostatico dopo intervento chirurgico oppure vigile attesa si potrebbe verificare quanto segue: la mortalità specifica risulta ridotta nel gruppo "intervento" rispetto al gruppo "vigile attesa", però la mortalità totale non è diversa tra i due bracci. Come mai? Una spiegazione può essere che lo studio non ha una potenza statistica tale da mettere in evidenza differenze sulla mortalità totale; un'altra, però, potrebbe essere che decessi che si sono verificati nel gruppo "chirurgico" in seguito a complicanze post-operatorie (per esempio embolie polmonari, infezioni, ecc.) sono stati classificati come "morti non dovute a cancro prostatico", il che porta ad una diminuzione della mortalità specifica ma non di quella totale.
E' per questo che alcuni esperti sostengono che sarebbe più corretto, in molti studi (come per esempio quelli sugli screening oncologici), valutare l'end-point MORTALITA' TOTALE E NON LA MORTALITA' SPECIFICA da CANCRO. Vediamo questo esempio, volutamente paradossale, che mi serve solo per spiegare meglio questo punto. In uno studio su uno screening oncologico succede che nel gruppo randomizzato allo screening la neoplasia viene scoperta molto più precocemente rispetto al gruppo "non screening", il tumore è in uno stadio operabile e quindi i pazienti sono avviati all'intervento chirurgico; al contrario nell'altro braccio i tumori vengono diagnosticati così tardivamente da essere quasi sempre inoperabili, quindi quasi nessuno si sottopone all'intervento. Al termine dello studio si trova che i decessi attribuiti al tumore sono stati decisamente inferiori nel gruppo screenato, ma i decessi totali sono stati maggiori per una elevata percentuale di complicanze dovute all'intervento chirurgico. Se lo studio valutasse solo l'end-point "decessi da cancro" non darebbe una informazione corretta perchè in realtà la mortalità totale sarebbe maggiore nel gruppo screenato.

End-pointsecondari

Veniamo adesso ad un aspetto particolare che è quello degli **end-pointsecondari** , croce e delizia degli esperti di “critical appraisal”, cioè di quei tizi un po’ rompiscatole e bastian contrari che si divertono a fare le pulci ai trials.
Quando si studia il disegno di un trial si predefinisce un **end-pointprimario** che ci si propone di valutare alla fine dello studio, singolo o composto. Conoscere qual è l'end-point primario di un trial è di capitale importanza perché è quello sul quale viene tarata la potenza statistica, cioè in pratica è quello che definisce lo scopo dello studio stesso. E' diventato, però, prassi comune definire anche uno o più **end-pointsecondari** , singoli o composti. Non è detto che l'end-point primario sia più importante di quello secondario, dal punto di vista clinico, ma è quello sul quale si dovrebbe basare l'interpretazione statistica del trial. Gli end-point secondari sono utili quando vanno nella stessa direzione di quelli primari, ma se per esempio l'end-point primario non risulta ridotto in maniera significativa mentre quello secondario lo è, può essere arduo trarre delle conclusioni affidabili e convincenti. Infatti, lo ripetiamo, è sull'outcome primario che si sono svolti i calcoli preliminari di tipo statistico.
Ma gli autori, nelle loro conclusioni, tengono sempre conto di questo "caveat"?
Purtroppo non è così.
Per esempio in uno studio vengono paragonati due trattamenti antipertensivi.
L'end-point primario sia costituito dall'infarto non fatale. Alla fine dello studio tale end-point non differisce statisticamente tra i due gruppi, ma si registra una riduzione di end-point secondari (per esempio stroke, eventi cardiovascolari totali, mortalità totale) in uno dei due gruppi. Gli autori concludono che un regime antipertensivo è più efficace di quello di paragone, ma non rilevano il fatto che l'end-point primario era uguale nei due gruppi. Una conclusione più corretta avrebbe invece dovuto richiamare il fatto che i due trattamenti sono equivalenti per l'end-point primario mentre la riduzione di alcuni end-point secondari richiede conferme da ulteriori studi.
Non si vuol qui sostenere che gli end-point secondari non siano importanti: possono di per sè essere pienamente validi ma andrebbero interpretati con più **cautela** perché una valutazione rigorosa dello studio, dal punto di vista statistico, dovrebbe sempre fare riferimento all'end-point primario. In altre parole una riduzione "statisticamente significativa" di un outcome secondario può esserlo solo in via nominale, cioè la sua rilevanza statistica può essere solo **apparente**. Infatti il potere "matematico" dello studio si riflette unicamente sull'outcome primario. I risultati derivanti dagli end-point secondari possono fornire informazioni supplementari rispetto a quelle trovate con l’end-point primario se sono concordanti. In caso contrario dovrebbero essere considerati soprattutto come IPOTESI da valutare in uno studio successivo.Insomma un'interpretazione del trial basata solo sugli end-point secondari andrebbe sempre guardata con prudenza, indipendentemente da quanto prestiosa sia la rivista che pubblica il lavoro.

Conclusioni

In conclusione, per chiudere sugli end-point, mi sento di dare i seguenti consigli:
1. Quando qualcuno presenta uno studio magnificando le virtù del tal farmaco chiedere sempre quali erano gli end-point considerati dai ricercatori e abituarsi a dubitare se si tratta di end-point surrogati. Dubitare però non vuol dire non credere perché nulla vieta che un farmaco che ha dalla sua solo una dimostrazione di efficacia su end-point surrogati non possa, in futuro, disporre di studi che ne dimostrino l'utilità anche su outcomes clinici importanti. In genere il problema riguarda farmaci immessi in commercio recentemente, essi sono ancora troppo nuovi per avere già a loro merito studi con esiti clinici che richiedono molti pazienti arruolati e vari anni di follow-up. In questi casi è utile **sospendere il giudizio** , come facevano certi filosofi del buon tempo antico, e aspettare l'arrivo di studi più affidabili. Nel frattempo conviene usare farmaci alternativi più vecchi(di solito il mercato è sovrabbondante) e sperimentati e di cui si conosce anche meglio, proprio perché da più tempo in uso, il profilo di sicurezza a lungo termine. Questa strategia permette di evitare, con una certa ragionevolezza, di incorrere in effetti collaterali non noti, che sono più spesso prerogativa dei farmaci più recenti e usati da minor tempo.
2. valutare sempre con occhio critico la superiorità di un trattamento rispetto ad un altro se questo giudizio si basa solo su end-point secondari.

RenatoRossi