Pros and cons degli end points combinati

Data 08 agosto 2010
Categoria scienze_varie

Gli end points combinati presentano vantaggi, ma si prestano anche a problemi interpretrativi che possono addirittura inficiare le conclusioni dello studio.

Al fine di valutare l'importanza nella scelta degli end points compositi verrà analizzato uno studio recensito da Renato Rossi (http://www.pillole.org/public/aspnuke/news.asp?id=4570) che verrà brevemente ricapitolato.

Sono stati arruolati 390 pazienti affetti da diabete tipo 2 in trattamento con insulina randomizzati all'aggiunta di placebo oppure di metformina (850 mg 1-3 volte al giorno). Il follow-up è stato di 4,3 anni e l'end-point primario era un outcome composto da **mortalità e morbilità micro e macrovascolari** . End-point secondari erano la mortalità e la morbilità micro e macrovascolari valutate separatamente.
Il trattamento con metformina è riuscito a prevenire l'aumento di peso, ha migliorato il controllo glicemico ed ha ridotto il fabbisogno di insulina. Non ha però migliorato l'end-point primario.
Tuttavia la metformina risultava associata ad una riduzione degli eventi macrovascolari (Hazard Ratio 0,61; 95%CI 0,40-0,94; p = 0,02), in parte spiegabile con la riduzione del peso corporeo. Il numero di soggetti che è necessario trattare per evitare un evento macrovascolare è stato di 16,1.
Gli autori concludono che i loro risultati giustificano l'uso della metformina (a meno che non sia controindicata) nel diabetico tipo 2 dopo l'introduzione dell'insulina.

Fonte:

Kooy A et al. Long-term Effects of Metformin on Metabolism and Microvascular and Macrovascular Disease in Patients With Type 2 Diabetes Mellitus. Arch Intern Med. 2009 Mar 23;169:616-625.

Nel presente commento verrà preso in considerazione solo un aspetto del lavoro di Adrian e Kooy in quanto vincolante per la credibilità dei risultati complessivi e esemplificativo degli effetti della scelta del tipo di end point composito. Sarà pertanto analizzata in dettaglio solo l'analisi riferita all'end-point primario. I risultati primari (ossia quelli su cui è stata programmata la numerosità del campione) sono in fatti gli unici in grado di dare una risposta al quesito della ricerca:

diabetici di tipo 2 trattati con metformina l'incidenza dell' end-point primario è diversa rispetto all' incidenza riscontrata nei diabetici non trattati con metformina?

Ogni risultato secondario (ossia non programmato nel calcolo della numerosità campionaria) dovrebbe infatti essere utilizzato solo a sostegna dei risultati primari o -al limite- solo per generare ipotesi di lavoro . L'outcome primario è in questo caso rappresentato da un end-point composito (tabella A) per il quale gli autori riportano nella tabella 4 del lavoro originale il numero di eventi registrati in 4.3 anni di follow-up per 14 sottocomponenti.

L'utilizzo di un end-point primario rappresentato da un outcome composito è molto comune nella ricerca clinica.

Vantaggi degli end-point compositi

I vantaggi associati all' utilizzo di un end-point composito sono prima di tutto caratterizzati da un aumento dell' efficienza statistica del trial. Utilizzando un outcome composito la frequenza basale dell' evento risulta di solito maggiore rispetto a quella che caratterizza ogni suo sottocomponente. Nello studio di Kooy infatti il Rischio Basale complessivo, riportato dagli autori solo come percentuale, è pari a 0.28: ciò significa che nel gruppo assegnato a placebo il 28% dei pazienti ha subito l'end-point composito durante il follow-up considerato.
Rispettivamente, la frequenza dell'end-point nel gruppo di intervento è risultata pari a 0.31 (il 31% dei pazienti assegnati a metformina ha subito l'end-point, con una differenza cruda in rischio assoluto pari a ARR=0.03, ossia pari a 3 punti percentuali). Si vede bene dalla tabella A che il rischio basale dell'end-point primario è maggiore rispetto al rischio basale di qualsiasi suo sottocomponente. Un incremento del rischio basale si esprime in termini statistici nella possibilità di

a) utilizzare una numerosità campionaria inferiore rispetto a quella che si dovrebbe utilizzare analizzando un singolo sottocomponente
b) utilizzare periodi di follow-up più brevi
c) aumentare la potenza statistica dello studio, ossia la capacità di riconoscere differenze quando queste sono reali.
Come valore aggiunto l'utilizzo di un end-point composito può fornire maggiori informazioni rispetto a quelle fornite da un singolo sottocomponente: per esempio, se un intervento è in grado di ridurre gli IM non fatali, ma aumenta gli IM

fatali la sua pericolosità potrebbe essere rilevata utilizzando l'outcome composito (IM fatali + IM non fatali), ma non utilizzando l'outcome singolo (IM non fatali).

Svantaggi degli end-point compositi

Tuttavia occorre utilizzare un end-point composito in modo ragionato e prudente perché le insidie legate a questo tipo di analisi sono molte .
In primo luogo in base al **Principio della Coerenza** i sottocomponenti dell'end-point composito non dovrebbero essere né troppo simili, né troppo diversi.
L' utilizzo di sottocomponenti troppo simili è caratterizzato da ridondanza dell'informazione (una eccessiva interdipendenza tra i vari sottocomponenti non apporta alcun valore aggiunto all' interpretazione generale del disease) e da una riduzione del potenziale incremento del rischio basale dell' evento (vedi sopra) che costituisce il miglior valore aggiunto associato all' utilizzo di questi end-point.
Allo stesso tempo i sottocomponenti dell'outcome composito non dovrebbero essere troppo dissimili. Questa situazione è esattamente antitetica a quella configurata in precedenza. Una "totale indipendenza" offre vantaggi per il sample size perchè consente di ottenere i valori massimi di rischio basale dell'evento . Una "notevole indipendenza" sottende però grandi differenze tra i sottocomponenti ed è quindi spesso associata a dubbi interpretativi. A parte qualche eccezione, più i sottoelementi che compongono un outcome composito sono differenti, minore è la visione "di insieme" del processo patologico di cui l'end-point dovrebbe essere espressione (che rappresenta –lo si ricordi- la motivazione più importante della scelta di un end-point composito.
La scelta ideale è rappresentata dall' aggregare in uno stesso end-point sottoelementi caratterizzati da un certo grado di dipendenza in quanto associati ad un comune processo fisiopatologici, ma allo stesso tempo sufficientemente diversi da poter fornire individualmente informazioni sul decorso della malattia a cui non avrebbero potuto contribuire gli altri end-point. Molto probabilmente nella ricerca di Kooy & coll esiste ridondanza nelle informazioni fornite dai singoli sottocomponenti.
Ricaviamo infatti dalla tabella A che il numero di pazienti che hanno subito l'outcome composito corrisponde a 55 soggetti nel braccio di controllo e a 65 soggetti nel braccio di intervento. Questi numeri non sono riportati dal testo, come invece si sarebbe dovuto fare, ma sono calcolati da chi scrive dalle percentuali crude dell' evento indicate a pag 620 dell'articolo nei 194 pazienti assegnati a placebo e nei 196 pazienti assegnati a metformina (0.28 x 194 = 54.32 0.31x196 = 64.76).
Per cui, complessivamente, 115 pazienti hanno subito nel corso del follow-up l'end-point primario (54.32+64.76=115.08).
Ricordiamo che il numero di pazienti che hanno subito l' outcome primario risponde al principio della "first occurrence", ossia viene riportato per la costruzione dell'end-point composito solo un sottocomponente per ogni paziente, quello subito cioè per primo. Significa per esempio che se un paziente prima ha subito un intervento coronarico e poi un infarto miocardico contribuisce alla costruzione dell'outcome composito "solo" per il sottocomponente "cardiovascular intervention".
Sottraendo dal numero di eventi riferiti ai sottocomponenti complessivamente registrati nel corso della ricerca il numero di soggetti che hanno subito l'end-point composito possiamo stimare in modo indiretto il grado di interdipendenza tra i varioutcome.
Infatti qualora tutti i pazienti avessero subito solo un evento il numero dei soggetti che hanno subito l'end-point composito coinciderebbe perfettamente con la somma degli eventi che riguardano i sottocomponenti dell' end-point composito: la loro differenza sarebbe cioè uguale a zero. Se la somma degli eventi complessivamente registrata per tutti i sottocomponenti dell' end-point composito è maggiore del numero di pazienti che hanno subito l' end-point composito significa ovviamente che un numero variabile di pazienti ha subito nel corso della ricerca più di un sotto-endpoint.
La differenza tra questi due dati esprime il numero di eventi riferiti ai sottocomponenti che riguardano pazienti che hanno subito più di un evento. Nella ricerca in oggetto il 66.1% dei 340 eventi che si riferiscono al numero dei sottocomponenti dell' end-point composito si riferisce a pazienti che hanno subito più di un evento.
Esiste pertanto con ogni probabilità una notevole ridondanza delle informazioni fornite da ogni singleton end-point sull' andamento generale della condizione clinica oggetto della ricerca.
Questa considerazione è importante in quanto qualora una scelta più oculata dei sottocomponenti della ricerca di Kooy avesse garantito un rischio basale dell' end-point composito più elevato, la potenza statistica dello studio avrebbe potuto essere sufficiente a dimostrare (cosa che non è avvenuta) la "significatività" della differenza rilevata tra i due bracci a confronto.
Un altro principio che dovrebbe condizionare le modalità con cui si costruisce un end-point composito è il **Principio della Rilevanza** .
Significa in pratica che i sottocomponenti dell'end-point dovrebbero essere outcome importanti per la salute del paziente. Se per esempio in un outcome costruito su 4 end-point uno solo dei sottocomponenti rappresentasse un obiettivo di salute importante (es : mortalità) e gli altri 3 rappresentassero outcome surrogati (es: normalizzazione della glicemia, della pressione, della colesterolemia) eventuali risultati "non significativi" o , nell' ipotesi peggiore, "dannosi" sulla mortalità potrebbero essere mascherati dai risultati associati agli outcome surrogati. Nella ricerca di Kooy questo principio è stato rispettato, in quanto solo il 7.2 % degli eventi riferiti ai sottocomponenti dell' end-point primario si riferisce ad un outcome surrogato (Ischemic changes of ECG)
La costruzione di un outcome composito dovrebbe obbedire anche al principio della **Omogeneità dell'Effetto** . Vale a dire che non dovrebbero esistere in primo luogo importanti differenze nei rischi basali dell' evento associati ai singoli sottocomponenti. In secondo luogo non dovrebbero esistere importanti differenze nei rischi relativi degli evento associati ai singoli sottocomponenti. Infine dovrebbe essere evitato l' impiego di sottocomponenti caratterizzati da outcome "clinician driver" ossia condizionabili da scelte soggettive potenzialmente diverse da parte di diversi operatori.
Il fatto che sia auspicabile che il rischio basale dei vari sottoelementi non sia troppo diverso è legato al fatto che a

parità di efficacia ciò costituisce un elemento a favore della omogeneità dell' effetto dell'intervento su tutto il pool. I "singleton end-point" clinicamente meno importanti dovrebbero pertanto essere caratterizzati da un livello di rischio basale paragonabile a quelli degli outcome clinicamente più rilevanti . Il rischio assoluto dei singoli end-point può essere ricavata dai dati che si riferiscono ai relativi gruppi di controllo. Nella ricerca di Kooy et al i rischi basali dell' evento oscillano per i vari sottocomponenti dell' end-point primario entro un range compreso tra 0.0% e 13.9% in 4.3 anni (media: 6.0%, ds 4.2%): è chiaro quindi che in tal senso il principio di omogeneità dell' effetto non è stato rispettato.
In secondo luogo, in uno studio di eventi gli end-point compositi offrono vantaggi in termini di efficienza statistica quando i risultati di efficacia rilevati per i singoli sottoelementi risultano tutti nella stessa direzione. Questa circostanza non è la regola e ciò è spesso imputabile ad una scelta incongrua dei sottoelementi.
L'omogeneità dell' effetto può essere valutata confrontando i rischi relativi (o qualsiasi altra misura di efficacia) calcolati per i vari sottoelementi.
Il riscontro di omogeneità è sempre a sostegno di una scelta oculata dei singleton end-point. La motivazione principale che giustifica la scelta di un end-point composito è infatti che i risultati ad esso riferiti rappresentino una proxy dell'andamento generale della malattia.
Combinare sottoelementi troppo diversi può condurre ad una sovrastima o a una sottostima dell' importanza clinica dell'end-point composito quando il risultato sia guidato dai sottoelementi meno importanti.
Nello studio di Kooy esiste incoerenza nella direzione e nell'intensità dell'effetto registrata nella frequenza di ogni singolo sottocomponente dell'end-point composito in quanto il rischio dell' evento risulta aumentato nel braccio assegnato a metformina per gli outcome 'Myocardial infarction', 'Peripheral arterial disease', 'Cardiovascular intervention', 'Non traumatic amputation', 'Sudden death' , 'Progression of retinopathy' (50% degli eventi che interessano complessivamente i sottocomponenti dell' outcome composito risultano così associati ad un aumento del rischio nel braccio di intervento). All'opposto, il rischio risulta ridotto per gli outcome 'Hearth failure' , 'Ischemic change of ecg' , 'Acute coronary syndrome', 'Diabetic foot' , 'Transient ischemic attack', ossia per il 25.3% complessivo degli eventi. Nessuna differenza tra i due bracci è emersa infine per gli outcome 'Stroke', 'Progression of nephropathy', 'Progression of neuroropaty', vale a dire per il 24.7% degli eventi che riguardano i sottocomponenti dell'outcome composito.
E' chiaro che più outcome si analizzano nello stesso campione, più i risultati ad essi attribuiti possono essere spiegati solo dal caso, come probabilmente è accaduto anche per questa analisi. Ma la direzione e l'intensità dell' effetto registrata per l' end-point primario è strettamente associata alla combinazione di questi risultati, per cui appare a chi scrive molto difficile interpretare in modo congruo i risultati complessivi riportati dagli autori.
Infine occorre considerare il problema degli end-point clinician-driven. Freemantle in una revisione di 167 articoli pubblicati dal 1997 al 2001 in 9 riviste in cui venivano utilizzati end-point compositi ha dimostrato che nel 43% dei casi essi includevano sottoelementi rappresentati non da veri esiti clinici, ma piuttosto da procedure mediche o chirurgiche utilizzate come proxyes di esiti clinici (rivascolarizzazione, valvuloplastica percutanea, ventilazione meccanica, ospedalizzazione, trapianto, shunt, uso di rescue therapy, istituzionalizzazione, inizio di nuova terapia antibiotica, terapia antishock, amputazione, dialisi, ossigenazione extracorporea). Freemantle sottolinea che i risultati di efficacia che si riferiscono a questi outcome possono essere soggetti a variabilità molto maggiori rispetto a quanto di solito si osserva per esiti più importanti come ad esempio l'incidenza di infarto o la mortalità. La decisione di sottoporre un paziente ad intervento chirurgico, ad esempio, può essere influenzata da opinioni personali del medico o da correnti di 'scuola' o di 'pensiero'.

Conclusioni

Gli autori riportano "non significatività statistica" per i risultati aggiustati associati alla terapia con metformina confrontata vesrus placebo .
Queste conclusioni dovrebbero però essere interpretate, a giudizio di chi scrive, con massima cautela.
In primis il numero di sottocomponenti scelti per l'end-point composito appare di per se eccessivo. L'errore random associato ai singoli risultati di efficacia può aver condizionato in questa situazione i risultati riferiti all' intero end-point composito. Per esempio se considerassimo il cut off di errore alfa accettabile a livello di ogni singolo confronto per giudicare "non dovuto al caso" il risultato rilevato in uno dei 14 sottocomponenti, in presenza di un livello medio di interdipendenza fra gli outcome pari al 66.1% e accettando un errore alfa complessivo pari a 0.05 il cut off necessario per stabilire significatività statistica per ciascuna analisi corrisponderebbe a $P<0.003571429$. All'opposto, la probabilità di riscontrare risultati spiegati solo dal caso a livello di uno o più confronti corrisponderebbe in questa situazione, ossia analizzando complessivamente 14 end-point, al 51.23%.
La diversità rilevate entro il pool delle 14 analisi nella intensità e , soprattutto, nella direzione dell'effetto testimonia indirettamente (dato che gli outcome sono tutti associati alla condizione clinica di diabete) che parte di questi risultati può essere solo effetto del caso.
In seconda istanza, oltre alla diversità rilevabile nei risultati riferiti ad ogni singolo sottocomponente anche l'eterogeneità dei rischi basali dell' evento rilevabile entro lo stesso pool testimonia uno scarso rispetto del principio di coerenza, in quanto a parità di dimensioni dell'effetto l'eterogeneità viene spiegata in tal caso anche dalla diversità nei rischi basali dell'evento.
In terza istanza, a parità di altri fattori, la ridondanza di informazioni offerte dall' utilizzo di sottocomponenti clinicamente simili può aver ridotto la frequenza basale dell' evento e la potenza statistica dell' analisi, contribuendo ad un eccesso di errore beta che potrebbe spiegare la "non significatività" dei risultati esposti.
In quarta istanza gli autori hanno calcolato il sample size prevedendo una quantità di errore beta eccessiva (a pagina 618 affermano infatti di aver calcolato il sample size su una potenza statistica pari a 75%, contro il cut off di 80% comunemente accettato) e basando l' analisi, senza motivazioni esplicite, su un test statistico a una sola coda.
Questi artifici, al pari dell'utilizzo di un end-point primario composto da 14 sottelementi, hanno contributo sicuramente alla programmazione di una numerosità campionaria non elevata e quindi ad una ottimizzazione delle risorse investite, ma non contribuiscono certamente all'affidabilità dei risultati riportati dal trial.

AlessandroBattaggia

Riferimentibibliografici

Moyé LA : Multiple Analyses in Clinical Trials Fundamentals for investigators. Springer 2003 ISBN 10918937

Freemantle N.: How well does the evidence on pioglitazone back up researchers' claims for a reduction in macrovascular events? BMJ 2005 331:836

Montori V et al. : Validity of composite end points in clinical trials. BMJ 2005 330:594