



Continuiamo a nascondere la zia scema in soffitta?

Alessandro Battaglia e Alberto Vaona



EQM (Evidenza, Qualità e Metodo in Medicina Generale)

Leggiamo in anteprima su pillole.org i risultati dello studio PROACTIVE [9], già egregiamente commentato da Renato Rossi negli interessanti aspetti clinici e a cui aggiungiamo solo considerazioni squisitamente metodologiche, giocoforza -per la loro stessa natura- un pò 'forti'. Le conclusioni degli autori rispecchiano infatti in modo paradigmatico quello che il dott. Moyè ha definito nel 1999 'nascondere la zia scema* in soffitta' [1].

Non abbiamo in mano l' articolo originale (che ci procureremo al più presto) e i nostri commenti si riferiscono quindi solo ai dati che si possono ricavare dall' abstract.

Il lavoro pubblicato sulla prestigiosa rivista The Lancet appare ad un sommario esame (appunto limitato all' abstract) un capolavoro di ingegneria metodologica architettato per orientare a favore dell' ipotesi preconstituita (= <il pioglitazone fa solo bene>) tutte le analisi consentite dal trial. Lo scopo per cui si organizza una ricerca è però più preciso: rispondere ad un quesito attraverso l'analisi dei risultati di efficacia rilevati per l' outcome primario. In questo caso il quesito era: il pioglitazone riduce la morbilità macrovascolare nei pazienti ad alto rischio con diabete 2 ?

La definizione di morbilità cardiovascolare era rappresentata dall' outcome composito < all-causes mortality, non fatal myocardial infarction (including silent myocardial infarction), stroke, acute coronary syndrome, endovascular or surgical intervention in the coronary or leg arteries, and amputation above the ankle>.

Gli autori hanno quindi generosamente mescolato ben sei outcome nella composizione del puzzle del loro end-point primario (alcuni dei quali a loro volta compositi) avendo come obiettivo - sicuramente raggiunto- quello di reclutare una popolazione ad alto rischio per questo end-point. Lo vediamo dal Rischio Assoluto dei controlli ($ARc = 572/2633 = 0.21$), da cui leggiamo che *in assenza dell' intervento un pò più di un paziente su cinque ha subito l' evento* entro il follow-up considerato. Leggiamo allo stesso modo che assumendo il farmaco *un pò meno di un paziente su cinque ha subito l' evento* entro il follow-up considerato ($ARi = 514/2605 = 0.19$). Arruolare <pazienti ad alto rischio> incrementa notevolmente l' efficienza statistica di un trial in quanto spesso amplifica eventuali differenze tra i due bracci e consente un aumento della precisione della stima che esprime queste differenze (5). Tutto ciò si traduce in maggiori probabilità di confermare l' efficacia di un trattamento (quando esiste) e, a parità di potenza statistica, di arruolare un numero minore di soggetti ad un determinato livello di errore alfa (di solito: cut off di $P=0.05$).

Nonostante questi encomiabili sforzi cosa è successo nello studio PROACTIVE?

E' successo che L' Hazard Ratio che esprime queste differenze tra i due bracci (lo si interpreti come un Rischio Relativo) è, sì, pari a 0.90 - e ciò equivale a dire che <il pioglitazone ha abbattuto del 10% il rischio assoluto riscontrato nei controlli> ma...i suoi dannati intervalli di confidenza erano compresi tra 0.80 e 1.02!

La adorata <significatività statistica> non è stata quindi, ahimè, raggiunta.

Ed è proprio un peccato per gli autori, direbbe sir Archibald Cochrane , in quanto la potenza di uno studio (leggasi: <la capacità di dimostrare differenze tra i due bracci quando effettivamente esistono> viene tarata PROPRIO sull' outcome primario, quello che condiziona ex ante la possibilità di rispondere al quesito della ricerca e su cui quindi viene stata tarata la numerosità del campione.

Tutto ciò a che scopo? dimostrare -qualora esista- una differenza tra i due bracci considerata 'ex ante' <cl clinicamente importante> dagli autori.

In questo caso quali avrebbero dovuto essere le conclusioni?

Queste: <non esistono dati a conferma di una differenza tra i due bracci in termini di 'macrovascular morbidity and mortality', in quanto abbiamo definito 'macrovascular morbidity and mortality' l' outcome primario <

< mortalità da tutte le cause + infarti non fatali (compresi quelli silenti) + stroke + sindrome coronarica acuta + interventi chirurgici vascolari o endoarteriectomie nell' albero coronarico o nelle arterie degli arti + amputazioni d' arto sopra la caviglia >.

Come spesso accade invece cosa hanno fatto gli autori del PROACTIVE?

hanno nascosto la zia scema in soffitta (zia scema = risultati sull' outcome primario) non commentando, almeno nell' abstract, quei miserabili e inestetici valori di P (per l' HR = 0.9 dell' outcome primario i valori di P erano 0.095, quindi >0.05).

Era più conveniente infatti abbracciare appieno i più gratificanti risultati rilevati per un outcome diverso da quello primario ossia per l' outcome secondario (anch'esso super-composito): <all-cause mortality, non-fatal myocardial infarction, and stroke>, dove la significatività statistica era stata raggiunta con P=0.027.

Le conclusioni riportate nell' abstract, bypassando allegramente l' ipotesi iniziale a cui il risultato rilevato per l' outcome primario -come in tutte le sperimentazioni correttamente interpretate - doveva fornire una risposta, sono che <il pioglitazone riduce la somma degli eventi 'mortalità generale', 'infarto non fatale' e 'stroke' in diabetici di tipo 2 ad alto rischio>.

Recenti esempi tratti dalla letteratura confermano che il rigore metodologico viene spesso e purtroppo immolato sull' altare delle personali opinioni.

I ricercatori basano troppo spesso le proprie conclusioni sulle analisi di end-point secondari e dei sottogruppi [3].

In particolare i risultati rilevati per gli outcome secondari vengono enfatizzati come alternativa alla <mancata dimostrazione di significatività statistica > dei risultati che riguardano l' outcome primario [2; 3].

La metodologia ci insegna invece che l' analisi degli end-point secondari (e dei sottogruppi) dovrebbe essere utilizzata solo come valore aggiunto a supporto delle conclusioni ricavate dall' analisi primaria [4] oppure - al limite- allo scopo di generare ipotesi di lavoro [3].

Nello studio ASCOT-BPLA per esempio -pubblicato assai di recente sulla stessa rivista [5]-, gli autori in base a) ai risultati rilevati per 16 outcome secondari per dieci dei quali la potenza statistica non era adeguata a dimostrare le differenze riscontrate - vedi tab. 1 - [6] e b) ai risultati dell' analisi di un end-point secondario eseguita su 18 sottogruppi del campione originale commentano quanto osservato raccomandando di <non utilizzare nella terapia antiipertensiva - ogni volta sia possibile- l' atenololo (eventualmente associato al tiazidico) al posto dell' amlodipina (eventualmente associata al pirendopril> nonostante ASCOT-BPLA non abbia dimostrato differenze significative tra i due bracci in termine di incidenza dell' outcome primario (= <mortalità coronarica + infarti non fatali>).

Tabella 1- Congruità del campione relativamente alle analisi degli outcome secondari* dello studio ASCOT-BPLA [5]

Outcome secondari nello studio ASCOT -BPLA	Amlodipina (n= 9639)		Atenololo (n=9618)		Potenza dello studio per la differenza riscontrata tra i due bracci nella frequenza di ciascun outcome*	
	n. eventi	Rischio Assoluto braccio di intervento	n. eventi	Rischio Assoluto braccio di controllo	Errore beta	%sample size ideale
Non fatal MI (excluding Silent IM) + Fatal CHD	390	0,0405	444	0,0462	0,51	48,1
Total Coronary Endpoint	753	0,0781	852	0,0886	0,25	88,5
Total Cardiovascular events and procedures	1362	0,1413	1602	0,1666	0,00	301,5
All Cause Mortality	738	0,0766	820	0,0853	0,40	62,4
Cardiovascular Mortality	263	0,0273	342	0,0356	0,09	141,6
Fatal and Non fatal Stroke	327	0,0339	422	0,0439	0,05	164,1
Fatal and Non Fatal Hearth Failure	134	0,0139	159	0,0165	0,69	27,7
Silent IM	42	0,0044	33	0,0034	0,80	15,8
Unstable Angina	73	0,0076	106	0,011	0,31	77
Chronic Stable Angina	205	0,0213	208	0,0216	0,97	0,3
Peripheral Arterial Disease	133	0,0138	202	0,021	0,03	186
Life-threatening arrhythmias	27	0,0028	25	0,0026	0,95	0,9
Development of Diabetes Mellitus	567	0,0588	799	0,0831	0,00	549,6
Development of Renal Impairment	403	0,0418	469	0,0488	0,35	69,5
Primary Endpoints + Coronary Rev. Proc.	596	0,0618	688	0,0715	0,23	92,8
Cardiovascular Death + Myocardial Infarction + Stroke	796	0,0826	937	0,0974	0,05	164,1

* per alfa = 0.05

Nello studio di Ridgker sulla prevenzione primaria cardiovascolare con ASA nel sesso femminile [7] -altro esempio- gli autori , pur riconoscendo che nelle donne esaminate l' ASA <non esercita alcuna influenza sul rischio di infarto miocardio o di morte cardiovascolare> sottolineano che < il rischio di stroke (outcome secondario) è però abbassato> e che < in un sottogruppo di donne con età superiore o uguale a 65 anni l'ASA è risultato efficace nel prevenire l'outcome primario, lo stroke ischemico e l'infarto miocardico>.

L'analisi per sottogruppi eseguita in questo lavoro riguardava 31 sottocategorie di pazienti entro ciascuna delle quali i due bracci erano stati messi a confronto per 4 outcome diversi: in tutto quindi erano stati eseguiti 124 confronti entro lo stesso campione ! Per 24 di questi confronti (19,3 per cento) i risultati erano accompagnati da livelli di significatività espressi da valori di P <0,05 . Per 100 (80,6 per cento) i risultati erano caratterizzati da valori di P maggiori o uguali a 0,05 ('non significativi'). In presenza di un numero così elevato di sottogruppi molti risultati potevano essere dovuti solo all'effetto del caso [3] . In più i confronti aspirina-placebo con risultati 'non significativi' (79 per cento dei sottogruppi con risultati 'statisticamente non significativi) potevano essere soggetti ad errore beta per inadeguata potenza statistica , spiegata dall' ovvio sottodimensionamento campionario entro l' ambito dei sottogruppi. Forzando la interpretazione del trial verso la direzione desiderata , gli autori valorizzavano inoltre solo uno tra i risultati 'significativi' di questa analisi: l' <efficacia del trattamento nelle donne più anziane>, attribuendo invece 'all'effetto del caso' un risultato a sostegno di una potenziale pericolosità dell' aspirina (l'aumentata paradossale incidenza dell'outcome primario nelle donne fumatrici) e ignorando risultati 'significativi' o 'borderline' di altri sottogruppi quando difficilmente spiegabili da un ragionamento fisiopatologico.

Conclusioni

Chi scrive sostiene, analogamente a più illustri opinioni, che esistano molti argomenti a favore di un utilizzo più prudente e ponderato delle analisi degli outcome secondari e dei sottogruppi [3].

Una volta scelto l' outcome primario tutta la potenza dello studio è stata spesa per le analisi ad esso riferite : l' effetto del caso può giocare quindi un ruolo molto forte nel condizionare i risultati di efficacia riscontrati in singoli sottogruppi o che riguardino outcome diversi da quello primario, anche in presenza della desiderata 'significatività statistica' [3].

La US Food and Drug Administration, coerentemente, per l' autorizzazione all' immissione in commercio di nuove molecole richiede trial randomizzati ben condotti in cui la potenza statistica sia tarata su differenze relative ad un outcome primario [8].

In un' epoca in cui qualsiasi Servizio Sanitario sempre più tende a considerare l' analisi costo-efficacia come parte integrante dei processi decisionali di rimborsabilità dei nuovi farmaci diventa sempre più importante che le evidenze scientifiche siano interpretate in modo corretto, al fine di evitare di incorporare nella pratica clinica scelte terapeutiche inappropriate [3]

Referenze bibliografiche

[1] Moyè LA

End-point interpretation in clinical trials: the case for discipline
Control Clin trials 1999 20:40-9

[2] Sackett DL

Why randomized controlled trials fail but needn't: 2. Failure to employ physiological statistics, or the only formula a clinician-trialist is ever likely to need (or understand)
JAMC 2001 165(9):1226-37

[3] Freemantle N

Interpreting the results of secondary end points and subgroup analyses in clinical trials: should we lock the crazy aunt in the attic?
BMJ 322 2001 322:989-91

[4] University of Washington

RCDRC Guidelines for Clinic Trial Designs
<http://depts.washington.edu/rcdrc/app.html>

[5] Björn Dahlöf, et al

'Prevention of cardiovascular events with an antihypertensive regimen of amlodipine adding perindopril as required versus atenolol adding bendroflumethiazide as required, in the Anglo-Scandinavian Cardiac Outcomes Trial-Blood Pressure Lowering Arm (ASCOT-BPLA): a multicentre randomised controlled trial'
Lancet 2005; 366: 895–906

[6] Battaglia A, Vaona A, Beradi S, Rigon G, Giustini SE

Old or new drugs for Hypertension? no problem: the result is the same [in press]

[7] Ridker PM et al

Randomized trial of low dose aspirin in the primary prevention of cardiovascular disease in women NEJM 2005:352A

[8] Fischer LD

Carvedilol and the Food and drug Administration (FDA) approval process: the FDA paradigm and reflections on hypothesis testing
Contr Clin Trials 1999 20:16-39

[9] John A Dormandy et al

Secondary prevention of macrovascular events in patients with type 2 diabetes in the PROactive Study (PROspective pioglitAzone Clinical Trial In macroVascular Events): a randomised controlled
The Lancet 2005 366:1279-1289