

# L'INTERAZIONE NEGLI STUDI EPIDEMIOLOGICI

**Graziella D'Arrigo, Fabio Provenzano, Carmine Zoccali,  
Giovanni Tripepi**

CNR-IBIM, Unità di Ricerca di Epidemiologia Clinica e Fisiopatologia delle Malattie Renali e dell'Iperensione Arteriosa, Reggio Calabria

## INTRODUZIONE

Negli articoli precedenti di questa serie abbiamo descritto come utilizzare la regressione lineare e la regressione logistica per testare ipotesi e controllare per il confondimento negli studi epidemiologici (1, 2). Attraverso queste due tecniche analitiche è possibile studiare l'interazione tra due fattori di rischio per spiegare/predire la frequenza e/o la severità di una specifica malattia o di un evento clinico nella popolazione. In termini generali, l'interazione tra un fattore di rischio A e un fattore di rischio B per spiegare la frequenza e/o la severità di una determinata patologia indica che il rapporto tra il fattore di rischio A e la gravità della malattia è dipendente dai livelli assunti dal fattore di rischio B (3). Cioè, il fattore di rischio B modifica il rapporto tra il fattore di rischio A e l'indicatore di risultato.

Immaginiamo di analizzare l'interazione tra il fumo e l'obesità per spiegare il rischio di infarto in un ipotetico campione di individui e di trovare una significativa interazione tra i due fattori di rischio. La presenza di questa interazione implica che il rapporto tra il fumo e l'infarto del miocardio è dipendente dal grado di obesità, cioè il rischio relativo (4) per l'infarto attribuibile al fumo è più alto negli obesi rispetto agli individui normopeso. In questo caso, l'obesità amplifica il rapporto tra il fumo e il rischio di infarto. È bene sottolineare che questa interazione va interpretata in termini bilaterali. In altri termini, anche il fumo modifica il rapporto tra l'obesità e il rischio di infarto, cioè il rischio relativo per l'infarto attribuibile all'obesità è più alto nei fumatori che nei non fumatori.

L'analisi dell'interazione riveste una particolare rilevanza nelle analisi secondarie dei *trial* clinici nei quali i ricercatori sono spesso interessati a sapere se il farmaco ha effetti diversi nel ridurre il rischio di un certo evento in specifici sottogruppi di pazienti (giovani *versus* anziani, diabetici *versus* non diabetici, uomini *versus* donne, ecc.). Negli esempi che seguono descriveremo l'applicazione dell'analisi dell'interazione a un contesto nel quale la variabile dipendente è continua (**Esempio 1**) e a uno nel quale la variabile dipendente è binaria (**Esempio 2**). I due esempi sono relativi a uno studio osservazionale, ma le tecniche descritte in questo articolo potranno essere applicate anche a studi che riguardano l'efficacia di interventi farmacologici e non farmacologici come i *trial* clinici.

## Come analizzare l'interazione

Il termine di interazione tra due variabili può essere un termine moltiplicativo (cioè il semplice prodotto tra i due fattori di rischio che si suppone interagiscano) (**Esempio 1**) oppure una variabile categorica che combina i gruppi in rapporto alla presenza/assenza dei due fattori di rischio (**Esempio 2**).

## KEY WORDS:

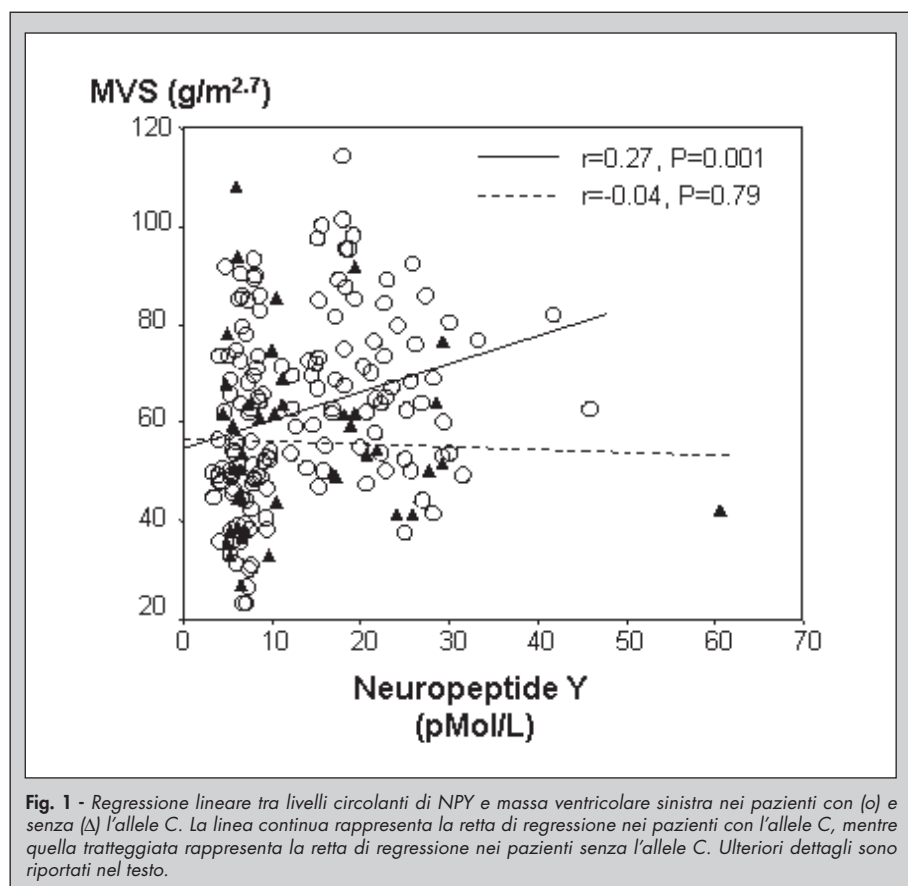
Epidemiology,  
Interaction,  
Linear  
regression,  
Logistic  
regression

## PAROLE CHIAVE:

Epidemiologia,  
Interazione,  
Regressione  
lineare,  
Regressione  
logistica

**ESEMPIO 1**

Il Neuropeptide Y (NPY) è un neurotrasmettitore che, attraverso recettori specifici (Y1-Y6), modula il rimodellamento vascolare e l'angiogenesi. I livelli plasmatici di NPY sono aumentati nei pazienti con insufficienza renale terminale e sono indipendentemente correlati all'ipertrofia ventricolare sinistra e agli eventi cardiovascolari in questi pazienti (5, 6). Poiché il polimorfismo (T585C) presente nell'esone 2 del gene NPY2R potrebbe influenzare i livelli circolanti di NPY, è stata analizzata l'interazione tra la concentrazione di NPY (7) e l'allele C di questo polimorfismo per spiegare la variabilità della massa ventricolare sinistra (MVS) in un gruppo di 189 pazienti in dialisi. Per testare questa ipotesi, gli Autori hanno preliminarmente disegnato la retta di regressione del rapporto tra i livelli circolanti di NPY e la MVS separatamente nei pazienti con e senza l'allele C (Fig. 1) e hanno trovato che il rapporto NPY-MVS è significativo nei pazienti con l'allele C ( $r=0.27$ ,  $P=0.001$ ) ma non in quelli senza questo allele ( $r=-0.04$ ,  $P=0.79$ ). Questa analisi evidenziava, inoltre, che il rapporto NPY-MVS aveva una pendenza maggiore nei pazienti con l'allele C che in quelli senza questo allele. Infatti, a uno stesso aumento di NPY (10 pMol/L) corrispondeva un incremento medio di 5.7 g/m<sup>2.7</sup> di MVS nei pazienti con l'allele C, mentre nessuna variazione si osservava nei pazienti senza questo allele (Fig. 1). L'aumento stimato di MVS corrispondente a un prefissato incremento dei livelli circolanti di NPY è stato calcolato utilizzando i coefficienti derivati dal modello di regressione lineare (1).



Per verificare se le due rette di regressione divergessero in maniera statisticamente significativa gli Autori hanno costruito un modello di regressione lineare che aveva come variabile dipendente la MVS e come variabili indipendenti:

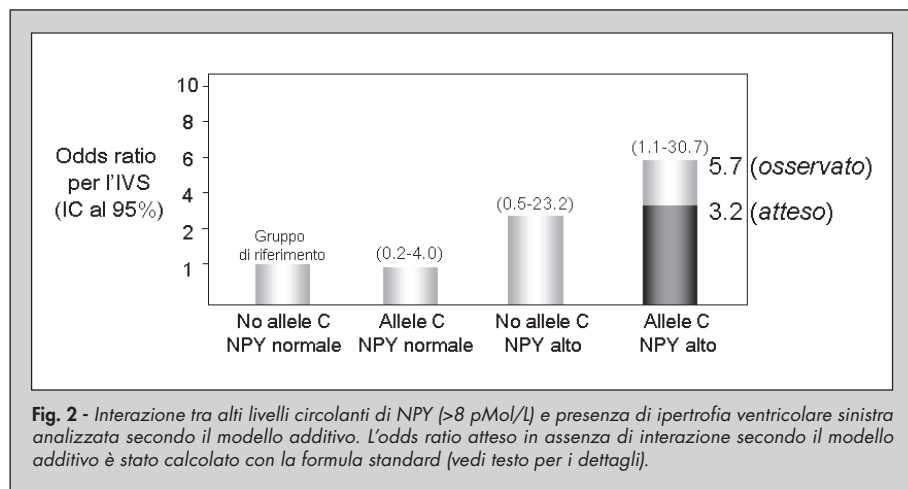
- i livelli circolanti di NPY (in pMol/L)
- il polimorfismo (in termini di variabile binaria, 0=assenza dell'allele C; 1=presenza dell'allele C)
- il termine moltiplicativo NPY\**polimorfismo* (o termine di interazione)

Quando si analizza l'interazione tra due fattori di rischio introducendo nel modello statistico il corrispondente termine moltiplicativo è importante introdurre nel modello anche i singoli fattori di rischio (nella fattispecie i livelli circolanti di NPY e il polimorfismo). Inoltre, l'aspetto più rilevante per il ricercatore è la significatività statistica del solo termine di interazione (la significatività statistica dei singoli fattori può essere ignorata). Nell'esempio sopra riportato, il termine di interazione NPY\**polimorfismo* è risultato statisticamente significativo ( $P=0.03$ ), indicando che la pendenza della retta che descrive il rapporto NPY-MVS è significativamente maggiore nei pazienti con l'allele C rispetto ai pazienti senza questo allele. Cioè, l'allele C modifica il rapporto tra livelli circolanti di NPY e MVS.

Per verificare se l'interazione NPY-allele C fosse indipendente da potenziali fattori di confondimento, gli Autori hanno fatto ricorso all'analisi multivariata. In un modello di regressione lineare multipla che includeva la MVS come variabile dipendente e i livelli circolanti di NPY, il polimorfismo, il termine di interazione NPY\**polimorfismo*, i fattori di rischio di Framingham, i fattori di rischio peculiari della malattia renale (età dialitica, emoglobina, albumina e il prodotto calcio-fosforo) e fattori di rischio emergenti (PCR e omocisteina) come variabili indipendenti, l'interazione NPY\**polimorfismo* diventava ancora più significativa ( $P=0.009$ ). Cioè, anche dopo aggiustamento per una serie di potenziali fattori di confondimento, la pendenza della retta che descrive il rapporto NPY-MVS è significativamente maggiore nei pazienti con l'allele C rispetto ai pazienti senza questo allele, confermando che l'allele C modifica l'effetto dei livelli circolanti di NPY sulla MVS indipendentemente da potenziali fattori di confondimento.

## ESEMPIO 2

L'interazione tra livelli circolanti di NPY e l'allele C per spiegare l'ipertrofia ventricolare sinistra (IVS) nei pazienti in dialisi è stata analizzata anche con la regressione logistica. I pazienti sono stati suddivisi in quattro gruppi sulla base della presenza/assenza dell'allele C e dei valori di NPY (normali/alterati, *cut-off* >8 pmol/L). In un modello di regressione logistica univariata, l'*odds ratio* (4) di IVS era più elevato nei pazienti con livelli alti di NPY e con l'allele C (*odds ratio*: 7.1, IC al 95%: 2.3-21.7), intermedio in quelli con NPY alto e senza l'allele C (*odds ratio*: 3.2, IC al 95%: 0.9-12.4) o con valori bassi di NPY e con l'allele C (*odds ratio*: 1.3, 95% CI: 0.5-3.7) e minimo in quelli con NPY normale e senza l'allele C (*reference group*, *odds ratio*=1) ( $P<0.001$ ). Questi risultati venivano confermati in un modello di regressione logistica multipla che includeva una serie di potenziali fattori di confondimento (Fig. 2). In questa analisi, l'*odds ratio* aggiustato dei pazienti con NPY alto e presenza dell'allele C era circa sei volte più alto (*odds ratio*: 5.7, IC al 95%: 1.1-30.7) di quello dei pazienti



senza questi fattori di rischio e maggiore di quello osservato nei pazienti con NPY elevato e senza l'allele C (odds ratio: 3.3, IC al 95%: 0.5-23.2) e nei pazienti con NPY normale e con l'allele C (odds ratio: 0.9, IC al 95%: 0.2-4.0). Per testare la presenza dell'interazione, quando i fattori di rischio sono combinati in sottogruppi, si utilizza la regola del modello additivo. Secondo questo modello, esiste un'interazione tra due fattori di rischio per spiegare un determinato indicatore di risultato se l'odds ratio osservato nei pazienti con entrambi i fattori di rischio è maggiore di quello atteso in assenza di interazione secondo il modello additivo. Nel nostro caso, l'odds ratio atteso si calcola sommando l'odds ratio per l'IVS dei pazienti con NPY alto e senza l'allele C (3.3) e quello dei pazienti con valori bassi di NPY e con l'allele C (odds ratio: 0.9) meno 1 (che rappresenta l'odds ratio del gruppo di riferimento), cioè:

$$\text{Odds ratio atteso} = 3.3 + 0.9 - 1 = 3.2$$

Poiché l'odds ratio osservato per l'IVS dei pazienti con NPY elevato e con allele C (5.7) era maggiore di quello atteso in assenza di interazione secondo il modello additivo (3.2), gli Autori dello studio hanno concluso che la presenza dell'allele C interagisce con gli elevati livelli di NPY per spiegare l'IVS nei pazienti in dialisi.

## CONCLUSIONI

L'analisi dell'interazione riveste un'importanza fondamentale negli studi epidemiologici in quanto consente di analizzare l'associazione tra uno specifico fattore di rischio e un determinato indicatore di risultato in rapporto ai livelli assunti da un'altra variabile denominata modificatore di effetto.

**Indirizzo degli Autori:**

Dr. Giovanni Tripepi, Statistician, MSc (Epidemiology)  
CNR-IBIM, Istituto di Biomedicina  
Epidemiologia Clinica e Fisiopatologia  
delle Malattie Renali e dell'Ipertensione Arteriosa  
Via Vallone Petrarà 55/57  
89124 Reggio Calabria  
e-mail: gtripepi@ibim.cnr.it

**DICHIARAZIONE DI CONFLITTO DI INTERESSI**

Gli Autori dichiarano di non avere conflitto di interessi.

---

**BIBLIOGRAFIA**

1. Provenzano F, Zoccali C, Tripepi G. Regressione lineare nella ricerca clinica. [Linear regression analysis in clinical research]. *G Ital Nefrol* 2011; 28 (1): 80-4.
2. Provenzano F, D'Arrigo G, Zoccali C, Tripepi G. La regressione logistica nella ricerca clinica. *G Ital Nefrol* 2011; 28: 210-3.
3. de Mutsert R, Jager KJ, Zoccali C, Dekker FW. The effect of joint exposures: examining the presence of interaction. *Kidney Int* 2009; 75 (7): 677-81.
4. Provenzano F, Tripepi G, Zoccali C. [How to measure effects in clinical research (corrected)]. *G Ital Nefrol* 2010; 27 (3): 296-300. Erratum in: *G Ital Nefrol* 2010; 27 (4): 422.
5. Zoccali C, Mallamaci F, Tripepi G, et al. Neuropeptide Y, left ventricular mass and function in patients with end stage renal disease. *J Hypertens* 2003; 21 (7): 1355-62.
6. Zoccali C, Mallamaci F, Tripepi G, et al. Prospective study of neuropeptide Y as an adverse cardiovascular risk factor in end-stage renal disease. *J Am Soc Nephrol* 2003; 14 (10): 2611-7.
7. Testa A, Mallamaci F, Macri R, et al. Neuropeptide Y receptor Y2 gene polymorphism interacts with plasma neuropeptide Y levels in predicting left ventricular hypertrophy in dialysis patients. *J Hypertens* 2010; 28 (8): 1745-51.