

LA REGRESSIONE DI COX

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

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

INTRODUZIONE

In un precedente articolo di questa serie abbiamo descritto il metodo di Kaplan-Meier, che permette di costruire due o più curve di sopravvivenza e di confrontarle tra loro con un particolare test di significatività statistica denominato *log-rank test* (1). L'analisi di Kaplan-Meier è il metodo d'elezione per confrontare le curve di sopravvivenza nei *trial* clinici controllati e randomizzati nei quali, per effetto della randomizzazione, trattati e non trattati risultano paragonabili per fattori di rischio noti e non noti (2, 3). Pertanto, ogni differenza osservata tra le curve di sopravvivenza potrà essere attribuita esclusivamente all'effetto del farmaco. Negli studi osservazionali, a differenza di quanto accade nei *trial* clinici, gli esposti ad un certo fattore di rischio possono differire dai non esposti per una serie di rilevanti caratteristiche cliniche (confonditori) (4), che possono alterare il rapporto tra l'esposizione oggetto dell'indagine da parte del ricercatore e l'incidenza di una specifica malattia o esito clinico. Il modello di Cox è una particolare tecnica di regressione multipla che permette di analizzare il rapporto tra un fattore di rischio (per esempio il fumo) e l'incidenza di un determinato esito clinico (per esempio l'infarto del miocardio), correggendo per uno o più fattori di confondimento (quali l'obesità e l'ipertensione).

KEY WORDS:

Hazard ratio,
Cox regression
analysis

PAROLE CHIAVE:

Hazard ratio,
Regressione
di Cox

LA REGRESSIONE DI COX

La regressione di Cox si usa negli studi di coorte, sia prospettici che retrospettivi. Nella regressione di Cox, la variabile dipendente è il tasso di incidenza di un determinato evento, cioè il numero di eventi per persona-tempo (5). Pertanto, a un determinato tempo t , per ogni individuo della coorte, è indispensabile conoscere la condizione (vivo/morto, evento/non evento, affetto/non affetto) e il tempo intercorso tra l'ingresso nello studio e la data in cui l'evento di interesse si è verificato (il cosiddetto tempo di sopravvivenza). Per gli individui che non hanno avuto l'evento oggetto dell'indagine epidemiologica, il tempo di sopravvivenza è il tempo intercorso tra l'ingresso nello studio e la data dell'ultima osservazione. Consideriamo uno studio di coorte che include un campione di 2500 pazienti in dialisi appartenenti al registro dell'ERA-EDTA. I pazienti sono stati arruolati tra il 1996 e il 2000 e sono stati seguiti prospetticamente fino al 31 Dicembre del 2005. L'indicatore di risultato dello studio è la morte. L'equazione generale di un modello di regressione di Cox avente l'obiettivo di analizzare il rapporto tra la presenza/assenza di un singolo fattore di rischio e un determinato esito clinico è la seguente:

$$H_t = H_{0t} \cdot \exp^{b \cdot x_i}$$

Dove " H_t " è il tasso di incidenza dell'evento (stimato dal modello) al tempo t , " H_{0t} " rappresenta il rischio di base (cioè il tasso di incidenza

dell'evento quando il fattore di rischio è assente), "b" è il coefficiente di regressione e "X_i" è il fattore di rischio.

Se calcoliamo il logaritmo naturale di entrambi i termini dell'equazione applicando le regole dei logaritmi, la regressione di Cox assume la forma:

$$\ln(H_i) = H_{0t} + b \cdot X_i$$

Il coefficiente di regressione (b) indica di quanto aumenta in media il logaritmo naturale del tasso di incidenza dell'evento negli esposti rispetto ai non esposti. Supponiamo di voler analizzare il rapporto tra la presenza/assenza del diabete e il tasso di incidenza della mortalità nella coorte dei 2500 pazienti in dialisi inclusi nel registro dell'ERA-EDTA.

Applicando l'equazione di Cox avremo:

$$\ln(H_i) = H_{0t} + b \cdot X_i$$

Dove:

- $\ln(H_i)$ rappresenta il logaritmo naturale del tasso di incidenza della mortalità negli individui della coorte;
- H_{0t} è il tasso di incidenza della mortalità nei non diabetici della coorte;
- b è il coefficiente di regressione che indica di quanto aumenta in media il logaritmo naturale del tasso di incidenza della mortalità nei diabetici rispetto ai non diabetici;
- X_i rappresenta il fattore di rischio (nella fattispecie il diabete, che ammette solo due possibilità: 0=assente, 1=presente).

Nei diabetici l'equazione di Cox sarà:

$$\ln(H_i) = H_{0t} + b \cdot 1 \quad (1 = \text{diabete presente})$$

Da cui

$$\ln(H_i) = H_{0t} + b$$

Nei non diabetici l'equazione di Cox sarà:

$$\ln(H_i) = H_{0t} + b \cdot 0 \quad (0 = \text{diabete assente})$$

Da cui

$$\ln(H_i) = H_{0t}$$

Per sapere di quanto aumenta in media il logaritmo naturale del tasso di incidenza della mortalità nei diabetici rispetto ai non diabetici, è sufficiente calcolare la differenza tra le due equazioni:

$$\ln(H_{\text{diabetici}}) - \ln(H_{\text{non diabetici}}) = H_{0t} + b - H_{0t}$$

Il termine H_{0t} ovviamente si annulla. L'equazione assume, perciò, la seguente forma:

$$\ln(H_{\text{diabetici}}) - \ln(H_{\text{non diabetici}}) = b$$

Per conoscere di quanto aumenta in media il tasso di incidenza dell'evento nei diabetici rispetto ai non diabetici è sufficiente calcolare l'esponenziale di entrambi i termini dell'equazione:

$$\exp^{\ln(H_{\text{diabetici}}) - \ln(H_{\text{non diabetici}})} = \exp^b$$

Poiché l'esponenziale della differenza tra due logaritmi naturali equivale a un rapporto, al primo termine dell'equazione avremo:

$$H_{\text{diabetici}} / H_{\text{non diabetici}} = \exp^b$$

L'esponenziale di **b** (cioè **2.7183^b**) rappresenta l'*hazard ratio*. Nella fattispecie, l'*hazard ratio* rappresenta il rischio relativo (6) ed indica quante volte è più alto il tasso di incidenza della mortalità nei diabetici rispetto ai non diabetici.

Nella Tabella I sono riportati i risultati della regressione univariata di Cox per quanto attiene il rapporto tra un singolo fattore di rischio (il diabete) e il tasso di incidenza della mortalità nel campione dei 2500 pazienti in dialisi appartenenti al registro dell'ERA-EDTA.

TABELLA I - ANALISI UNIVARIATA DI COX

Variabile	b	Errore Standard	P	Hazard ratio	Intervallo di confidenza 95%	
Diabete (sì/no)	0.54	0.07	<0.0001	1.71	1.5	1.9

Un coefficiente di regressione (**b**) pari a 0.54 indica che, nei diabetici, il logaritmo naturale del tasso di incidenza della mortalità è più alto di quello nei non diabetici di una quantità pari a 0.54. Calcolando l'esponenziale del coefficiente di regressione ($2.7183^{0.54}=1.71$) otteniamo l'*hazard ratio*. In questo caso, il tasso di incidenza della mortalità è del 71% più alto nei diabetici rispetto ai non diabetici. Il programma fornisce, inoltre, l'errore *standard* del coefficiente di regressione, la significatività statistica (P) e l'intervallo di confidenza al 95% dell'*hazard ratio* (6).

Per verificare se il rapporto tra diabete e tasso di incidenza della mortalità fosse indipendente dal potenziale effetto confondente dell'invecchiamento, gli Autori hanno costruito un modello di regressione multipla includendo l'età (Tab. II).

L'età può essere un potenziale confonditore, in quanto essa è associata sia al rischio di morte (l'indicatore di risultato) che all'esposizione (il diabete). Inoltre, l'età non è un effetto dell'esposizione e non vi sono evidenze che essa si trovi nella catena causale tra esposizione e rischio di mortalità nei pazienti in dialisi.

TABELLA II - REGRESSIONE MULTIPLA DI COX

Variabili	b	Errore Standard	P	Hazard ratio	Intervallo di confidenza 95%	
Età	0.047	0.002	<0.0001	1.048	1.043	1.053
Diabete (sì/no)	0.662	0.068	<0.0001	1.94	1.70	2.22

L'aggiustamento per l'effetto confondente dell'età influenza il rapporto tra diabete e tasso di incidenza della mortalità nei pazienti in dialisi, in

quanto l'*hazard ratio* aggiustato del diabete (1.94) è maggiore di quello crudo (1.71). L'età esercita un confondimento negativo, in quanto, dopo aggiustamento per età, l'*hazard ratio* di morte connesso al diabete aumenta di circa il 13%.

NUMERO DI COVARIATE NELLA REGRESSIONE DI COX

Il massimo numero di covariate che può essere inserito in un modello di regressione multipla di Cox è strettamente dipendente dal numero di eventi. Una semplice regola è quella di includere nel modello di regressione multipla 1 covariata ogni 10 eventi. Per esempio, se abbiamo un campione di 1000 individui che hanno 20 eventi durante il *follow-up*, il massimo numero di covariate da includere nel modello multiplo di Cox dovrebbe essere 2. Inoltre, per poter applicare la regressione di Cox, è indispensabile che i rischi siano proporzionali (7).

CONCLUSIONE

La regressione di Cox è un importante strumento statistico per valutare il rapporto tra esposizione ed indicatori clinici ed epidemiologici di risultato e per tenere sotto controllo il confondimento negli studi eziologici di coorte.

DICHIARAZIONE DI CONFLITTO DI INTERESSI

Gli Autori dichiarano di non avere conflitto di interessi.

Indirizzo degli Autori:

Dr. Giovanni Tripepi
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

BIBLIOGRAFIA

1. Tripepi G, Catalano F. [Kaplan-Meier analysis]. *G Ital Nefrol* 2004; 21 (6): 540-6.
2. Provenzano F, Tripepi G, Zoccali C. [Clinical trials (Part I)]. *G Ital Nefrol* 2010; 27 (4): 396-8.
3. Provenzano F, Tripepi G, Zoccali C. [Clinical trials (Part II)]. *G Ital Nefrol* 2010; 27 (5): 536-9.
4. Provenzano F, Versace MC, Tripepi R, Zoccali C, Tripepi G. [Confounding in epidemiology]. *G Ital Nefrol* 2010; 27 (6): 664-7.
5. Provenzano F, Tripepi G, Zoccali C. [Measures of frequency in epidemiology studies: incidence and prevalence]. *G Ital Nefrol* 2010; 27 (1): 78-81.
6. Provenzano F, Tripepi G, Zoccali C. [How to measure effects in clinical research (corrected)]. *G Ital Nefrol* 2010; 27 (3): 296-300.
7. van Dijk PC, Jager KJ, Zwinderman AH, Zoccali C, Dekker FW. The analysis of survival data in nephrology: basic concepts and methods of Cox regression. *Kidney Int* 2008; 74 (6): 705-9.