

Un modello generale per l'analisi della latenza in associazioni con esposizioni protratte e variabili nel tempo

Antonio Gasparrini

London School of Hygiene and Tropical Medicine



XXXVI Congresso dell'Associazione Italiana Epidemiologia
Bari, 29–31 Ottobre 2012

Il problema

Nella ricerca biomedica, è comune che eventi espositivi generino effetti ritardati nel tempo, un concetto qui definito **latenza**

Tale fenomeno è comune a vari **campi di ricerca**, tra cui:

- epidemiologia ambientale e occupazionale
- farmaco-epidemiologia
- studi life course

Problema: disegni di studio e modelli analitici diversi, dipendenti dai campi di ricerca



Un modello generale

Analiticamente e concettualmente, la difficoltà è nel modellare patterns di rischio per esposizioni **protratte e variabili nel tempo**

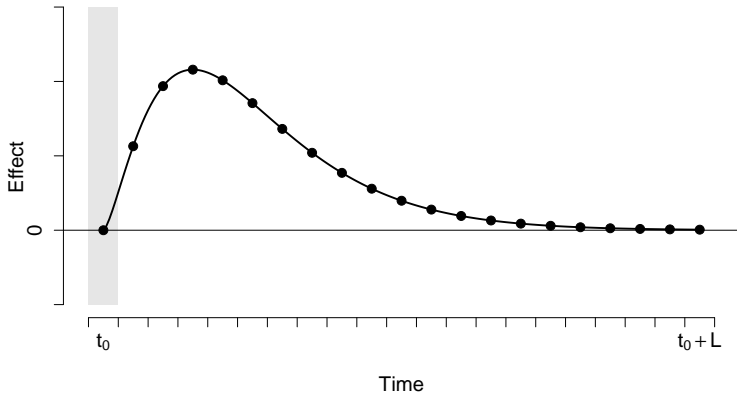
Il rischio può essere definito in termini di **intensità** e **timing** dell'esposizione \Rightarrow associazione **bi-dimensionale**

Scopo: sviluppare un framework concettuale e statistico per lo studio della latenza



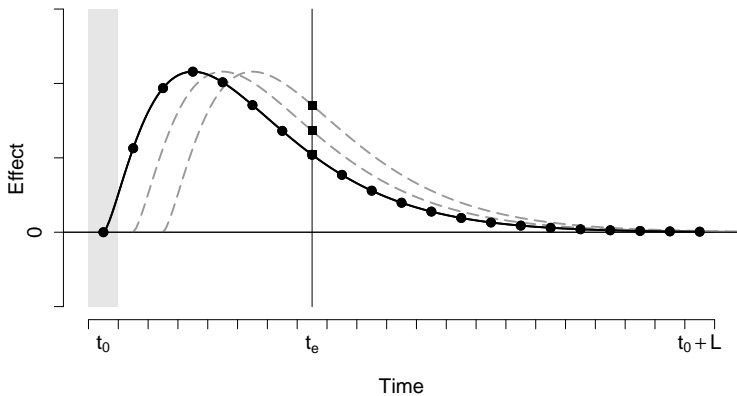
Modello concettuale

Singolo evento espositivo



Modello concettuale

Esposizioni multiple



Approcci in studi di coorte

Semplice **esposizione cumulativa** (Checkoway *AJIM* 1992)

Finestre espositive (Langholz *AJIM* 1999)

Approcci flessibili basati su una **funzione di pesi w** (Thomas *SJWEH* 1983, Hauptmann *Biometrics* 2000, Richardson *Epidemiol* 2009, Sylvestre *Stat Med* 2009)



Distributed lag models (DLMs)

Sviluppati per dati di serie temporale in econometria (Almon *Econometrica* 1965), e recentemente estesi all'epidemiologia (Schwartz *Epidemiology* 2000)

Basati sul concetto di **curva dei lag**, analogo alla curva dei pesi descritta prima, stimata da funzioni diverse (strati, polinomi, splines)

Framework analitico ben sviluppato, ma limitato a dati di serie temporali



Estensione non-lineare: DLNMs

Limitazione: assunzione di una relazione dose-risposta lineare

I DLM possono essere estesi per casi non-lineari, generando **distributed lag non-linear models (DLNMs)** (Armstrong *Epidemiology* 2006, Gasparrini *Stat Med* 2010)

Link con modelli precedenti proposti in altri campi (Abrahamowicz *Stat Med* 2007, Berhane *Stat Med* 2008)

Framework implementato e documentato nel **package dlrm** di R



Un esempio: la coorte CPUM

3347 minatori che lavorarono nel Colorado Plateau nel 1950–1960

Stato in vita e causa di decesso ottenuta da record linkage con follow-up al 1982 , con **258 morti per cancro del polmone**

Storie espositive a **radon** (WLM) e **fumo di tabacco** (pack×100) ricostruite for periodi quinquennali di età



Analisi

Associazione tra esposizione a radon e cancro del polmone

Modello di Cox per esposizioni **time-varying**

Età come asse temporale, controllo per fumo di tabacco e anno di calendario

Storie espositive in WLM/anno ricostruite per ogni soggetto/periodo entro **lag 2–40 anni**

DLNMs con funzioni differenti per stimare la **dose-riposta** e la **curva dei lag**: lineare, costante, strati, B-spline



Risultati da DLNM

Table: DLNMs con differenti funzioni per x e ℓ .

<i>Funzione di x</i>	<i>Funzione di ℓ</i>	<i>df</i>	<i>AIC</i>
Lineare	Costante	1	2581.5
Lineare	Strati ^a	4	2582.7
Lineare	B-spline ^b	3	2583.2
B-spline ^c	Costante	3	2528.0
Strati ^d	Strati ^a	8	2507.0
B-spline ^c	B-spline ^b	9	2500.1

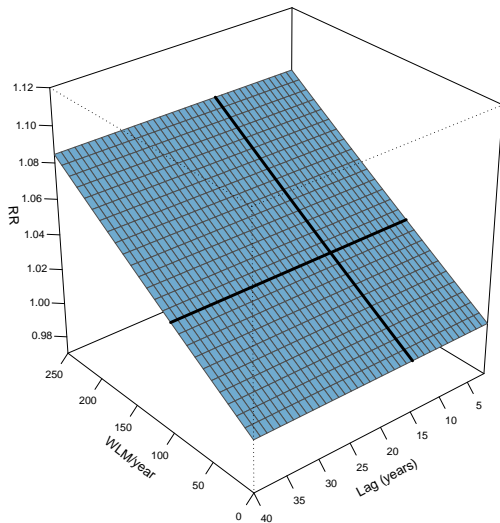
^a Cut-off a lag 10-20-30

^b Quadratica, 1 nodo a lag 13.3, senza intercetta

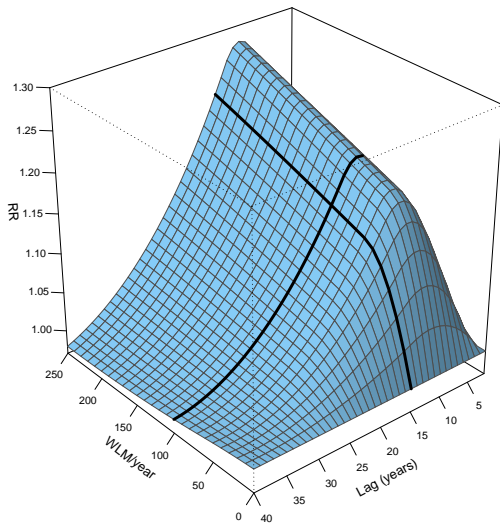
^c Quadratica, 1 nodo a 59.4 WLM/anno, senza intercetta

^d Cut-off a 26-121 WLM/anno

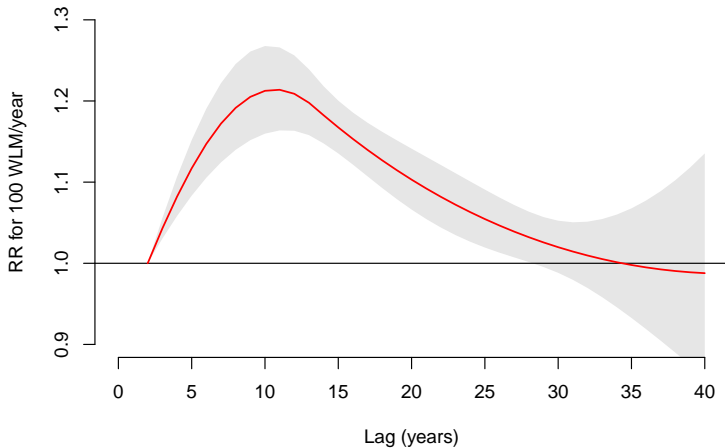
3D: esposizione cumulativa



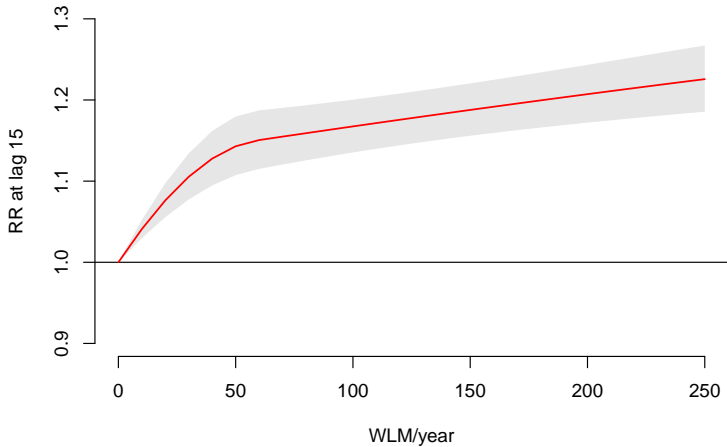
3D: B-spline e B-spline



Curva dei lag per 100 WLM/year

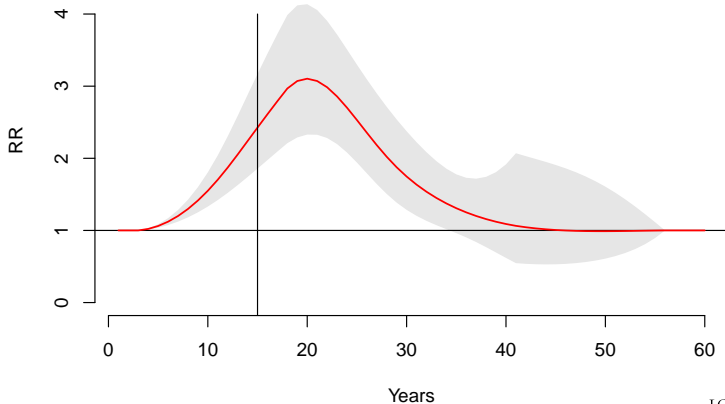


Dose-risposta a lag 15



Predizione nel tempo

20 WLM per 15 anni



Un modello unificato

I DLNM offrono un modello flessibile per esposizioni protrate e variabili nel tempo

Potenziali applicazioni in vari campi di ricerca, con **disegni di studio**, **dati** e **modelli di regressione** diversi

Framework basato su un modello concettuale che fornisce una **base interpretativa** e definisce chiaramente le **assunzioni**



Limitazioni

Il modello richiede la disponibilità di **storie espositive complete** (anche ricostruite)

L'esposizione deve essere **variabile nel tempo** entro soggetto

I modelli piu' complessi richiedono **statistical power** (qui 9df)

Problemi di **selezione del modello**



Bibliografia (I)

- Abrahamowicz et al (2006). Modeling cumulative dose and exposure duration provided insights regarding the associations between benzodiazepines and injuries. *Journal of Clinical Epidemiology*, 59(4):393-403.
- Abrahamowicz & MacKenzie (2007). Joint estimation of time-dependent and non-linear effects of continuous covariates on survival. *Statistics in Medicine*, 26(2):392408.
- Almon S (1965). The distributed lag between capital appropriations and expenditures. *Econometrica*, 33(1):178-196.
- Armstrong (2006). Models for the relationship between ambient temperature and daily mortality. *Epidemiology*, 17(6): 624631.
- Berhane et al (2008). Using tensor product splines in modeling exposure-time-response relationships: application to the Colorado Plateau Uranium Miners cohort. *Statistics in Medicine*, 27(26):5484-96.
- Checkoway & Rice (1992). Time-weighted averages, peaks, and other indices of exposure in occupational epidemiology. *American Journal of Industrial Medicine*, 21(1):25-33.
- Elliott et al (2007). Long-term associations of outdoor air pollution with mortality in Great Britain. *Thorax*, 62(12):1088-94.
- Gasparrini et al (2010). Distributed lag non-linear models. *Statistics in Medicine*, 29(21):2224-2234.
- Gasparrini (2011). Distributed lag linear and non-linear models in R: the package dlnm. *Journal of Statistical Software*, 43(8):1-20.
- Gasparrini (2012). Modelling of temporal dependencies in associations with time-varying exposures. *Statistics in Medicine*, in submission.
- Goodman et al (2004). Cause-specific mortality and the extended effects of particulate pollution and temperature exposure. *Environmental Health Perspectives*, 112(2):179185.



Bibliografia (II)

- Hauptmann et al (2000). Analysis of exposure-time-response relationships using a spline weight function. *Biometrics*, 56(4):1105-8.
- Langholz et al (1999). Latency analysis in epidemiologic studies of occupational exposures: application to the Colorado Plateau uranium miners cohort. *American Journal of Industrial Medicine*, 35(3):246-56.
- Langholz & Goldstein (1996). Risk set sampling in epidemiologic cohort studies. *Statistical Science*, 11(1):35-53
- Leffondre et al (2002). Modeling smoking history: a comparison of different approaches. *American Journal of Epidemiology*, 156(9):813.
- Richardson (2009). Latency models for analyses of protracted exposures. *Epidemiology*, 20:395399.
- Richardson et al (2009). Ionizing radiation and lymphoma mortality. *American Journal of Epidemiology*, 169(8):969-76.
- Roberts et al (2007). Maternal residence near agricultural pesticide applications and autism spectrum disorders among children in the California Central Valley. *Environmental Health Perspectives*, 115(10):1482-9.
- Schwartz (2000). The distributed lag between air pollution and daily deaths. *Epidemiology*, 11(3):320-326.
- Sylvestre & Abrahamowicz (2009). Flexible modeling of the cumulative effects of time-dependent exposures on the hazard. *Statistics in Medicine*, 28(27):3437-53.
- Thomas (1983). Statistical methods for analyzing effects of temporal patterns of exposure on cancer risks. *Scand J Work Environ Health* 1983;9(4):353-366.
- Thomas (1988). Models for exposure-time-response relationships with applications to cancer epidemiology. *Annual Review of Public Health*, 9:451-82.



Prospettive e assunzioni

Prospettive:

- **forward**: da un evento espositivo a una serie di rischi futuri
- **backward**: da un rischio a una serie di esposizioni passate

Assunzioni

- assunzione di **indipendenza**
- assunzione di **effetti identici**



Modelli con funzione di pesi

Una funzione $s(x, t)$:

$$s(x, t) = \beta_T \int_{t_0}^t w(t-u) x(u) du \approx \beta_T \sum_{u=t_0}^t w(t-u) x(u) \quad (1)$$

Re-parameterizzando in termini di ℓ :

$$s(x, t) = \beta_T \int_0^L w(\ell) x(t-\ell) d\ell \approx \beta_T \sum_{\ell=0}^L w(\ell) x(t-\ell) \quad (2)$$



Vantaggi statistici e computazionali

Lo sviluppo **teorico** e **statistico** dei DLNM esteso oltre le serie temporali

Basato interamente su modelli parametrici, con CI e test disponibili

I modelli possono essere fittati con funzioni di regressione standard

Framework **interamente implementato** nel package `dlnm` di R, disponibile sul CRAN (Gasparrini *JSS* 2011)



Il package dlnm di R

Esempio di codice

```
library(dlnm)

cb <- crossbasis(Q,lag=c(2,40),
  argvar=list(type="bs",degree=2,knots=59.4,cen=0),
  arglag=list(type="bs",degree=2,knots=13.3,int=F))

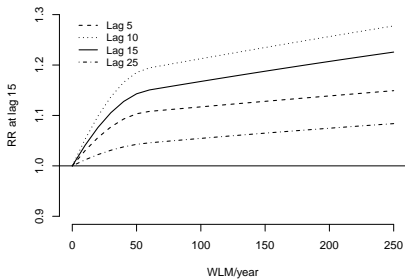
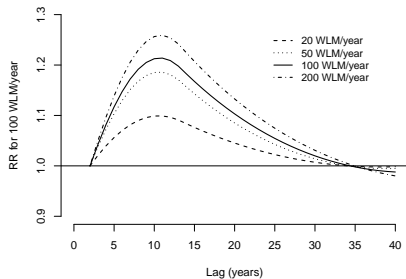
model <- coxph(Surv(agest,ageexit,ind)~cb+smoke+caltime,data)

pred <- crosspred(cb,model11,at=0:25*10)

plot(pred,"3d",xlab="WLM/year",ylab="Lag (years)",zlab="RR")
plot(pred,var=100,xlab="Lag (years)",ylab="RR")
plot(pred,lag=15,xlab="WLM/years",ylab="RR")
```



Slices



Rischio predetto

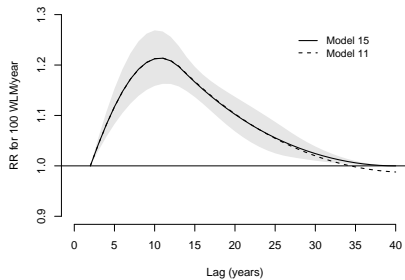
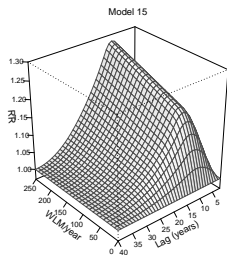
20 WLM per 10 anni

Table: Rischio predetto da modelli diversi

<i>Model</i>	<i>RR</i>
Lineare e costante	1.053 (1.045–1.062)
Lineare e B-spline	1.039 (1.027–1.050)
B-spline e costante	1.361 (1.245–1.487)
B-spline e B-spline	1.550 (1.331–1.805)



Constraints



Interazione

