A modelling framework for exposure-lag-response associations

Antonio Gasparrini

Medical Statistics Department
London School of Hygiene and Tropical Medicine

MRC Conference on Biostatistics - 25 March 2014
Health effects associated with exposures to risk factors always occur with some **delay**.

This issue encompasses study designs and statistical models. Some examples:

- Short-term exposure to air pollution and CVD risk
- Tobacco smoke and lung cancer risk
- Prescription of a drug and side effects
Previous research

Standard statistical approaches do not directly characterize this temporal structure

**Challenge:** modelling (potentially complex) temporal patterns of risk due to time-varying exposures

More advanced models previously proposed in cancer epidemiology (Duncan Thomas, Michael Hauptmann, David Richardson) and pharmaco-epidemiology (Michel Abrahamowicz)
Conceptual representation

Single exposure event

Forward perspective

<table>
<thead>
<tr>
<th>Time</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>t</td>
<td>●</td>
</tr>
<tr>
<td>t+1</td>
<td>●</td>
</tr>
<tr>
<td>t+2</td>
<td>●</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>t+L</td>
<td>●</td>
</tr>
</tbody>
</table>

0
Conceptual representation
New lag dimension

Time (Lags)
Effect
t₀
tₑ
t₀ + L
0
lag

Forward
Backward

L

0

Effect

Time (Lags)
Exposure-lag-response associations

The risk is represented by a function $s(x, t)$ defined in terms of both intensity and timing of a series of past exposures, expressed through:

- an exposure-response function $f(x)$ for exposure $x$
- a lag-response function $w(\ell)$ for lag $\ell$

Generating a bi-dimensional exposure-lag-response function:

$$s(x, t) = \int_{\ell_0}^{L} f \cdot w(x_{t-\ell}, \ell) \, d\ell \approx \sum_{\ell=\ell_0}^{L} f \cdot w(x_{t-\ell}, \ell)$$
Distributed lag models (DLMs)

Given a exposure history at time $t$ for lags $\ell = \ell_0, \ldots, L$:

$$ q_{x,t} = [x_{t-\ell_0}, \ldots, x_{t-\ell}, \ldots, x_{t-L}]^T $$

and assuming a linear exposure-response, we can write:

$$ s(x, t; \eta) = q_{x,t}^T C \eta = w_{x,t}^T \eta $$

where $C$ is obtained from the lag vector $\ell = [\ell_0, \ldots, \ell, \ldots, L]^T$ by applying a specific basis transformation.
Distributed lag non-linear models (DLNMs)

First the matrix $R_{x,t}$ is obtained applying a second basis transformation to $q_{x,t}$.

Then we define a tensor product:

$$A_{x,t} = (1_{v_{\ell}} \otimes R_{x,t}) \otimes (C \otimes 1_{v_{x}})$$

which forms the crossbasis:

$$s(x, t; \eta) = (1_{v_{x}}^{T} A_{x,t}) \eta = w_{x,t}^{T} \eta$$

The problem reduces to choosing a basis for each $q_{x,t}$ and $\ell$, defining exposure-response and lag-response functions, respectively.
An example

Aim: to study the association between radon exposure and mortality for lung cancer

3,347 subjects working in the Colorado Plateau mines between 1950–1960, 258 lung cancer deaths

Exposure history to radon (WLM) reported for 5-year age periods

Analysis with Cox proportional hazards model using age as time axis, controlling for smoking and calendar year
3D: linear-by-constant
3D: step-by-step
3D: spline-by-spline
Lag-response curves from DLNMs
Lag-responses at different exposures

![Graph showing lag-responses at different exposures. The x-axis represents lag in years, ranging from 0 to 40, and the y-axis represents RR, ranging from 0.9 to 1.3. Three lines represent different exposures: 20 WLM/year (dashed), 100 WLM/year (solid red), and 200 WLM/year (dashed blue).]
Exposure-responses at different lags

![Graph showing exposure-responses at different lags. The x-axis represents WLM/year, ranging from 0 to 250. The y-axis represents RR at lag 15, ranging from 0.9 to 1.3. The graph includes curves for different lags: Lag 5 (dashed black), Lag 15 (red), and Lag 25 (dotted blue).]
Predicted risk

Table: Predicted risk from different exposure histories

<table>
<thead>
<tr>
<th>Exposure history</th>
<th>HR</th>
</tr>
</thead>
<tbody>
<tr>
<td>20 WLM/yr in the last 10 yrs</td>
<td>1.52 (1.31–1.76)</td>
</tr>
<tr>
<td>100 WLM/yr in the last 10 yrs</td>
<td>2.37 (1.87–2.99)</td>
</tr>
<tr>
<td>20 WLM/yr in the last 20 yrs</td>
<td>3.12 (2.29–4.24)</td>
</tr>
<tr>
<td>20 WLM/yr 10–19 yrs ago</td>
<td>2.05 (1.70–2.48)</td>
</tr>
<tr>
<td>20 WLM/yr 30–39 yrs ago</td>
<td>1.04 (0.64–1.70)</td>
</tr>
</tbody>
</table>
Dynamic prediction: 20WLM for 15yrs
Statistical advantages

DLNMs offer a flexible way to model exposure-lag-response associations.

Unified framework based on a general conceptual and statistical definition, applicable in various study designs.

Based on entirely parametric approaches, with estimation and inferential tools of the shelf.

Models can be fitted with standard regression routines.
The framework is **fully implemented** in the R package `dlnm`, available from the CRAN (Gasparrini *JSS* 2011)

The package contains a **new vignette** focusing on applications beyond time series data
library(dlnm)

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

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

pred <- crosspred(cb, model, 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")
Simulations

Linear–Constant

Plateau–Decay

Exponential–Peak

Exposure

Lag

HR

0 1 2 3 4 5

0 10 20 30 40

1.0 1.02 1.04 1.06 1.08 1.10

True AIC avg AIC samples

Exposure

Lag

HR

0 1 2 3 4 5

0 10 20 30 40

1.0 1.02 1.04 1.06 1.08 1.10

True AIC avg AIC samples

Exposure

Lag

HR

0 1 2 3 4 5

0 10 20 30 40

1.0 1.02 1.04 1.06 1.08 1.10

True AIC avg AIC samples

Exposure

Lag

HR

0 1 2 3 4 5

0 10 20 30 40

1.0 1.02 1.04 1.06 1.08 1.10

True AIC avg AIC samples
Main references


E-mail: antonio.gasparrini@lshtm.ac.uk