Analysis of time-stratified case-crossover studies in environmental epidemiology using **R**

Aurelio Tobías

Spanish Council for Scientific Research

Nagasaki University School of Tropical Medicine and Global Health

XIV Summer School UPC-FME

Barcelona, 2021/07/02

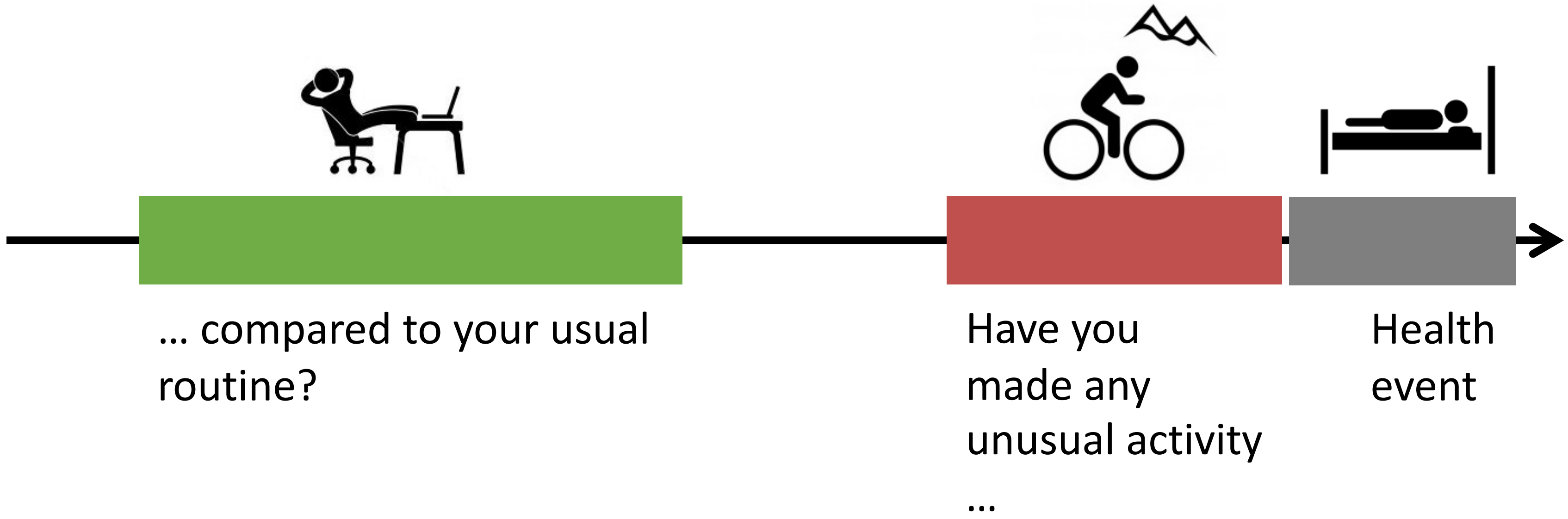
Outline

- Introduction
- The case-crossover design
- Conditional logistic regression
- Poisson regression
- Conditional Poisson regression
- Summary

Introduction

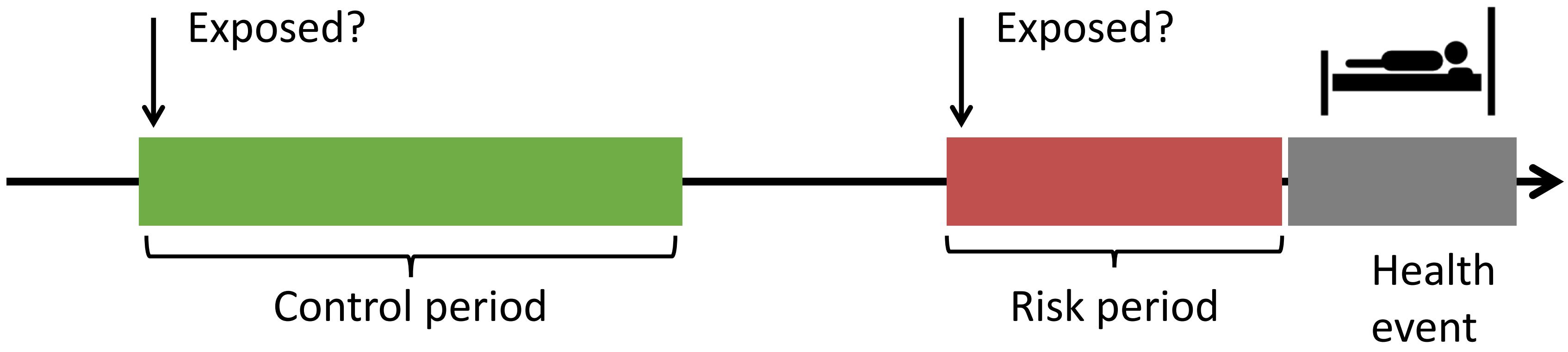
- Handling with individual factors it is not a straight forward process in conventional time series regression
- The time stratified case-crossover design is a popular alternative to for analysing associations between time series of environmental exposures and counts of health outcomes
- It was firstly proposed by *Maclure (1991)* to study transient effects on the risk of acute health events

The case-crossover design



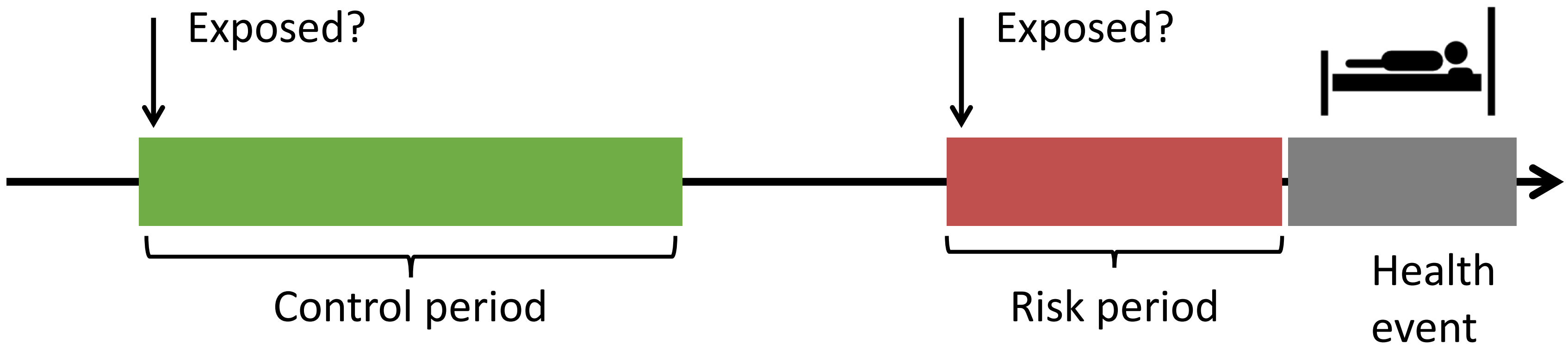
- Case only study (i.e., cases serve as their own controls)
- Comparison of exposure during *cases' index interval* versus exposure during *reference interval*

The case-crossover design



- Physical exercise and acute myocardial infraction *(Maclure 1991)*
- Use of mobile phone and car accident *(Redelmeier et al. 1997)*

The case-crossover design



- Analysis likewise a matched case-control

		Control	
		Exp.	No
Risk period	Exp.	a	b
	No	c	d

OR = b/c

The case-crossover design

Risk period	Control period (24 h)			Control period (1 week)	
	Exposed	Non-exposed		Exposed	Non-exposed
Exposed	4	22		5	27
Non-exposed	6	216		6	288
	OR = 3.7			**OR = 4.5**	

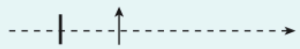
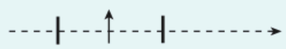
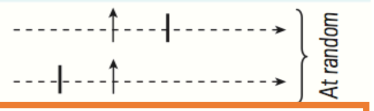
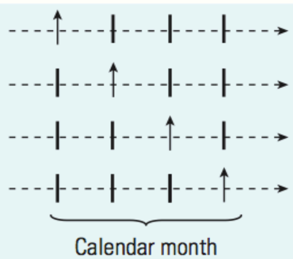
(Adapted from McEvoy et al. 2005 and Dawes 2005)

Application to environmental studies

- Firstly adapted to air pollution studies in Philadelphia *(Neas et al. 1999)* and Barcelona *(Sunyer et al. 2000)*
- Comparing exposure levels for a given day (t) when health event occurs *vs.* levels before (t-7) and after (t+7) the health event
- Allows to control for time-trend and seasonality by design, since it compares exposure levels between the same weekdays within each month of each year

Time-stratified case-crossover

Table 1. Comparison of different CCO designs.

Reference	Type	Selection of controls	Advantages	Factors that can introduce bias	Selection of controls diagram
Maclure 1991	CCO	One control point before the effect	All possible confounding factors undergoing no change between control periods and effect, automatically controlled for by design	Long-term trends or seasonality	
Bateson and Schwartz 1999	Symmetric bidirectional	Two at equal distance of the event	Provides adequate control for long-term trends and seasonality		
Navidi and Weinhandl 2002	Semisymmetric bidirectional	One chosen at random from the two used for symmetric bidirectional CCO	Provides adequate control for long-term trends and seasonality		
Lumley and Levy 2000	Time stratified	One (or several) within the same time stratum in which the event occurred	Provides adequate control for long-term trends and seasonality		

Arrows pointing up indicate case periods; horizontal arrows represent direction of time within 1 month; dashed lines indicate time periods of 1 day; vertical lines indicate control periods.

(Adapted from Carracedo et al. 2010)

Conditional logistic regression

- Dataset must be reshaped from time-series to individual matched case-control and analyzed with the **`clogit`** function in **R**

time-series

date	dow	deaths	pm
2001-01-01	mon	2	79.2
2001-01-02	...		
...			

individual matched

id	case	date	dow	pm
1	yes	2001-01-01	mon	79.2
1	no	2001-01-08	mon	49.2
1	no	2001-01-15	mon	80.1
1	no	2001-01-22	mon	60.4
1	no	2001-01-29	mon	64.1
2	yes	2001-01-01	mon	79.2
2	no	2001-01-08	mon	49.2
2	no	2001-01-15	mon	80.1
2	no	2001-01-22	mon	60.4
2	no	2001-01-29	mon	64.1

- Dataset reshaped from **n** days to **N × 4** individuals

Conditional logistic regression

- Similar approach to a matched case-control study familiar to the epidemiologist
- **But…**
    - Unfriendly data management
    - Very large data-sets at individual level
    - Slow convergence CPU time with interaction terms for individual variables
    - Can not account for overdispersion

Poisson regression

- Lu and Zeger (2007)
  - *" … the time-stratified case-crossover design leads to the same estimate as obtained from a Poisson regression with dummy variables indicating the strata"*
- Strata defined by the 3-way interaction between year-month-day of week

Autocorrelation and overdispersion

- Time-stratified model

$$Y_t = \alpha + \sum\beta_i year_i + \sum\delta_j month_j$$

- Periodic functions

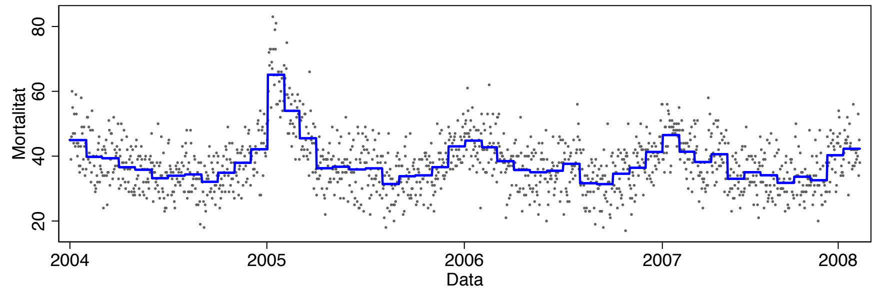
$$Y_t = \alpha + \beta t + \sum\delta_k \sin(^{k\pi t}/_T) + \sum\gamma_k \cos(^{k\pi t}/_T)$$

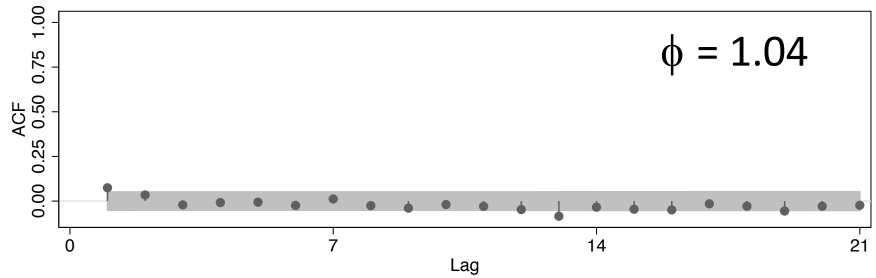
- Spline function

$$Y = \alpha + f(t_i, \beta)$$

- **Case-crossover**

$$Y = \alpha + \sum\beta_i year_i + \sum\delta_j month_j + \sum\delta_j dow_j + \sum\gamma_{ij} year_i \times month_j + \sum\gamma_{ij} year_i \times dow_j + \sum\gamma_{ij} month_i \times dow_j + \sum\gamma_{ijk} year_i \times month_j \times dow_k$$





Poisson regression

- Avoids reshaping the dataset to individual level
- Accounts for overdispersion with **`family=quasipoisson`**
- **But…**
  - Large number of interaction parameters, slowing the convergence CPU time, even making difficult with missing data
  - Much slower convergence CPU time with additional interactions for individual variables

Conditional Poisson regression

- Armstrong et al. 2014
  - *" … Poisson models with large numbers of indicator variables can alternatively be fit with conditional Poisson models, conditioning on numbers of events in the time stratum"*
  - Firstly proposed by *Farrington (1995)* for the self-controlled case-series design for vaccine safety studies
- Time-stratum set defined by grouping same weekday within each month of each year
- Use the **`gnm`** library in **R**

Conditional Poisson regression

date	dow	pm.	stratum
2001-01-01	**mon**	**79.2**	**1**
2001-01-02	tue	…	
2001-01-03	wed	…	
2001-01-04	thu	…	
2001-01-05	fri	…	
2001-01-06	sat	…	
2001-01-07	sun	…	
2001-01-08	**mon**	**49.2**	**1**
2001-01-09	tue	…	
2001-01-10	wed	…	
2001-01-11	thu	…	
2001-01-12	fri	…	
2001-01-13	sat	…	
2001-01-14	sun	…	
2001-01-15	**mon**	**80.1**	**1**
2001-01-16	tue	…	
2001-01-17	wed	…	
2001-01-18	thu	…	
2001-01-19	fri	…	
2001-01-20	sat	…	
2001-01-21	sun	…	
2001-01-22	**mon**	**60.4**	**1**
2001-01-23	tue	…	
2001-01-24	wed	…	
2001-01-01	thu	…	
2001-01-01	fri	…	
2001-01-01	sat	…	
2001-01-01	sun	…	
2001-01-29	**mon**	**64.1**	**1**

- From Poisson regression…

$$Y|x \sim \text{Poisson}(\mu)$$
$$E(Y|x) = \mu$$
$$\log(\mu) = \alpha + \beta x$$
$$\text{with } V(Y|x) = \mu$$

…to conditional Poisson regression

$$\mathbf{Y_{.,stratum}|x \sim \text{Poisson}(\mu)}$$
$$\mathbf{E(Y_{.,stratum}|x) = \mu_{.,stratum}}$$
$$\mathbf{\log(\mu_{.,stratum}) = \alpha_{stratum} + \beta x}$$
$$\text{with } \mathbf{V(Y|x) = \mu}$$

Conditional Poisson regression

- Avoids reshaping the dataset to individual level
- Accounts for overdispersion with **`family=quasipoisson`**
- Easy fit of interaction terms for individual variables with faster convergence CPU time
- In environmental epidemiology studies it is useful to
  - Analyse time-series with small-count data
  - Test for effect modification by individual characteristics (e.g., sex, age, neighborhood) and other environmental events (e.g., dust episodes, heatwaves)

Summary

- Time-stratified case-crossover accounts for temporal trends and short-term variations by design
- In environmental epidemiology studies can be used for small-count data and to easily test for effect modification
- Case-crossover studies can be analysed using **R**
  - Conditional logistic regression requires unfriendly data management, and large data-sets at individual level
  - Poisson regression with 3-way interaction requires to fit large number of interaction parameters, making it difficult convergence
  - **Conditional Poisson** solves the problems above allowing the analysis with faster convergence CPU time