Distributed lag nonlinear model

Chris Fook Sheng Ng

Department of Global Health Policy
Graduate School of Medicine
The University of Tokyo

Japan Science and Technology Agency

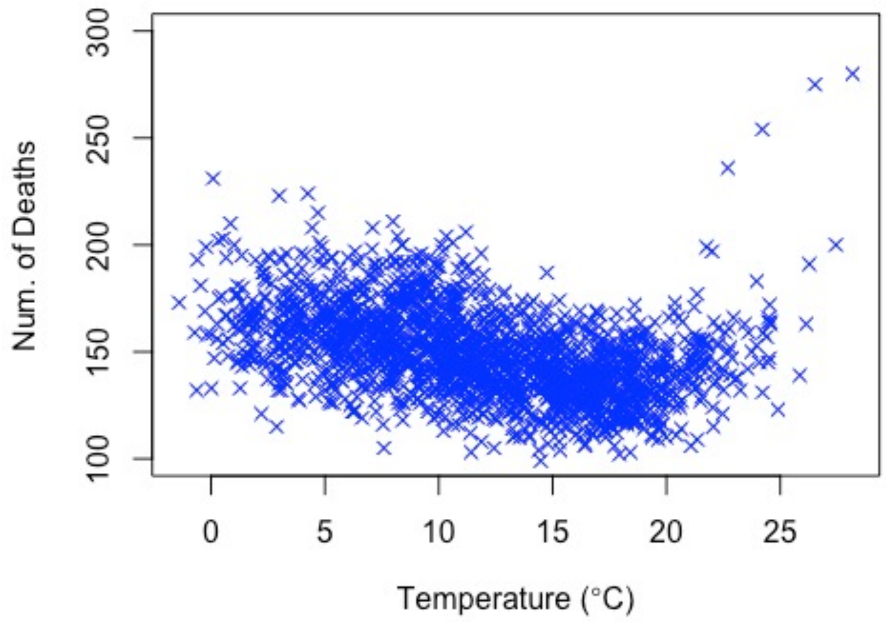
e-Asia Summer School 2023, Tokyo

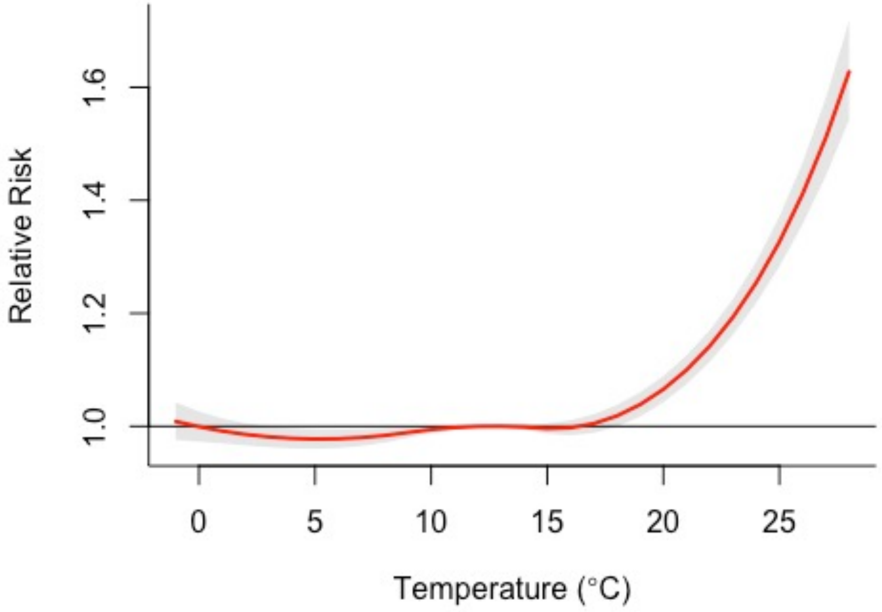
Outline

- Introduction
  - Non-linear association
  - What is a lagged effect?
- Distributed lags
  - Modelling strategies
  - Extension to nonlinear exposure-response association
- Common issues and suggestions

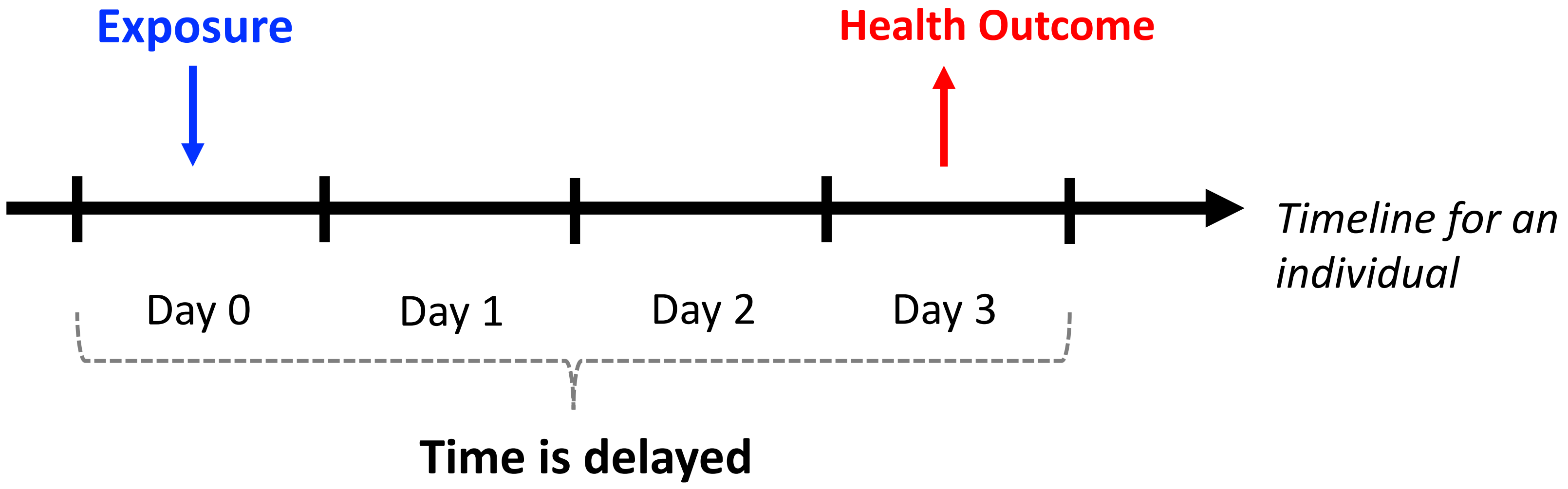
Refresher 1: non-linear association

- Non-linear exposure-response function



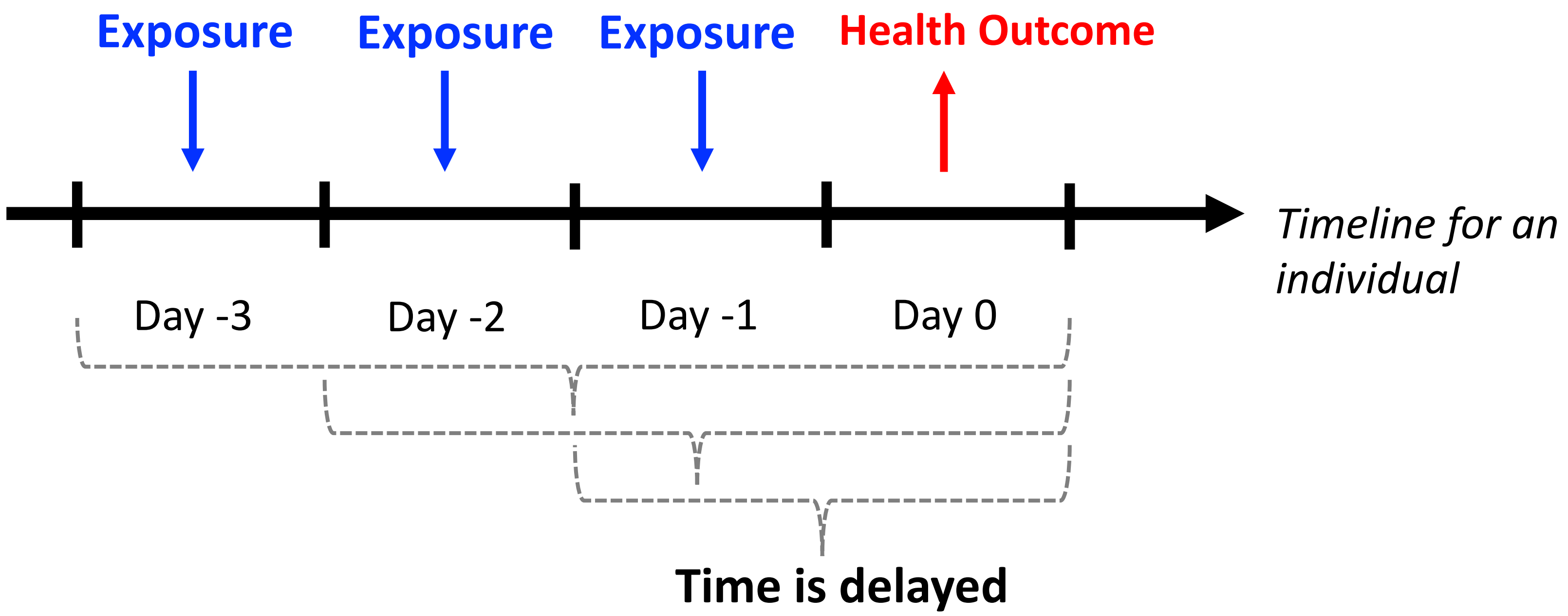


Refresher 2: lagged effect



- Conceptual figure shows time from exposure to health event is delayed
- Example for a single exposure

Multiple exposures and lagged effects



- Environmental exposure often continuous and time-varying (temperature/air pollution)
- Multilple lags with cumulative effects. Difficult to model
- Distributed lag model is useful

Model for lagged effects

Single-lag model

- One model one lag term*
- Each lag estimated independently
- Let x be exposure on time t, and cov represents a number of other factors such as day-of-week, a spline function for time, etc.

$$
\begin{aligned}
\log[E(Y_t)] &= \alpha + \beta_0 x_t + cov \\
\log[E(Y_t)] &= \alpha + \beta_1 x_{t-1} + cov \\
\log[E(Y_t)] &= \alpha + \beta_2 x_{t-2} + cov \\
&\vdots \\
\log[E(Y_t)] &= \alpha + \beta_{27} x_{t-27} + cov
\end{aligned}
$$

Distributed lag model

- All lags included in one model – distributed lags

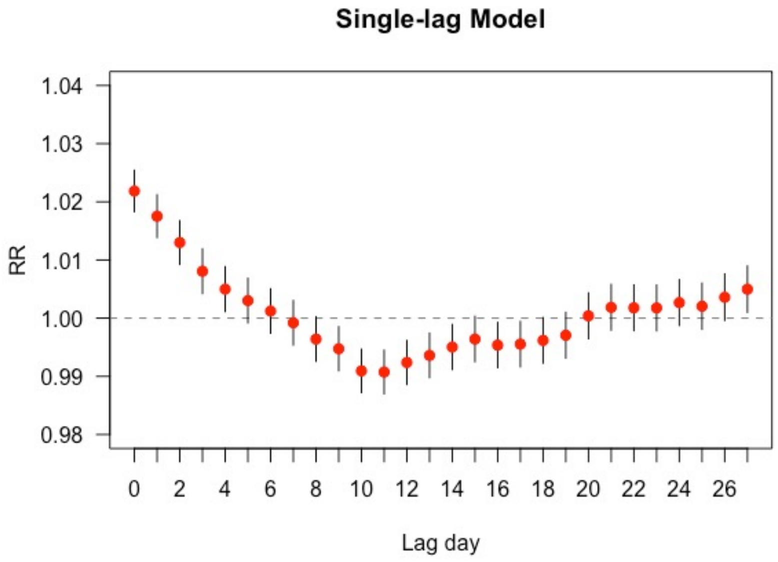
$$
\log[E(Y_t)] = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \cdots + \beta_{27} x_{t-27} + cov
$$

*Lag term = lag day = lag 0, 1, 2,… for exposure on the same day, 1 day before, 2 days before, etc.

Distributed lags

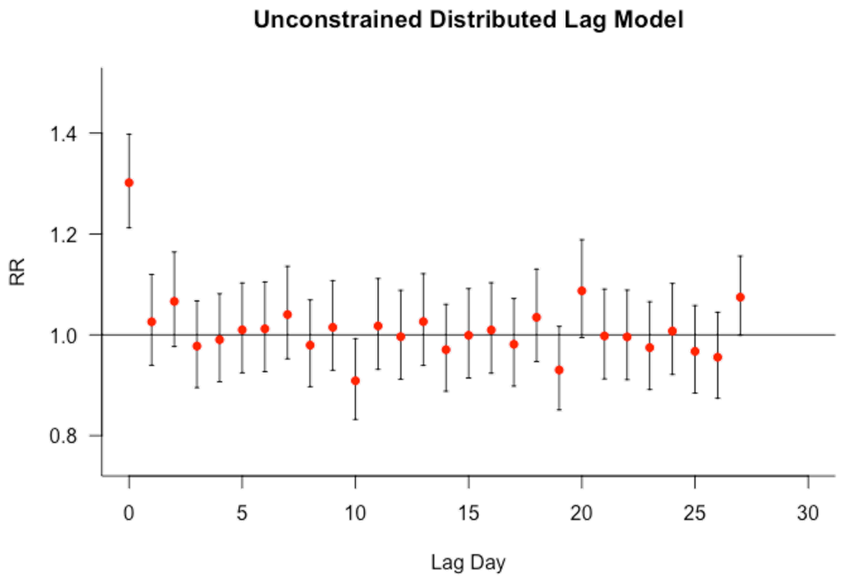
Single-lag model

- Lags are unadjusted for each other
- Difficult interpretations



Distributed lag model

- Highly correlated lags → larger error → wider confidence interval
- Noisy shape

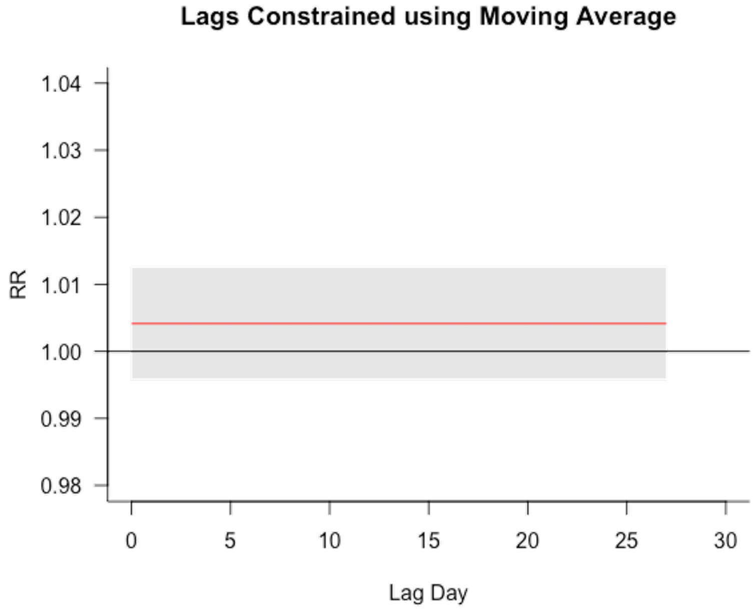


Summer heat and daily mortality, London 2002-2006 (June – September).
Temperature is linear, adjusted for ns(day-of-season, df=4)*year, and year.

Constraining the distributed lags

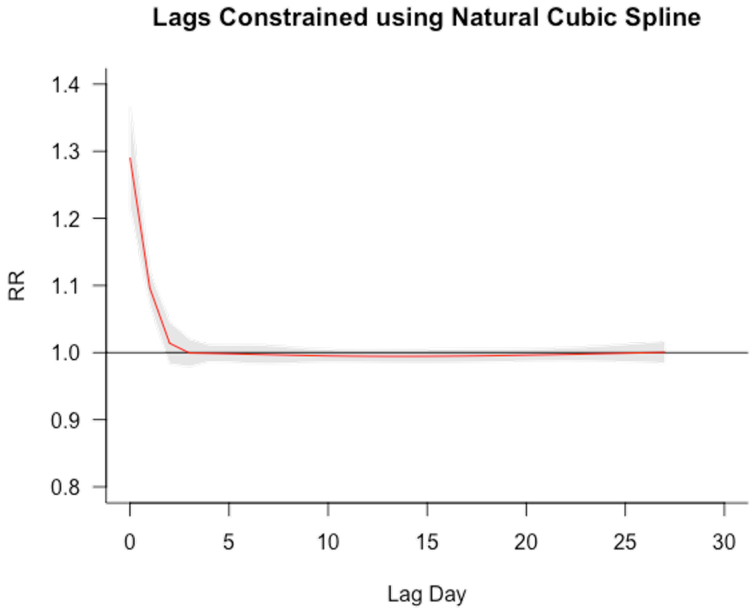
Using Moving Average

- Simple method
- Strong assumption: same effects for all lags
- No lag structure, overly simplified



Using Spline

- Smooth change (less noisy)
- Flexibility for nonlinear lag-response relationship
- eg. Natural cubic spline with 3 log knots

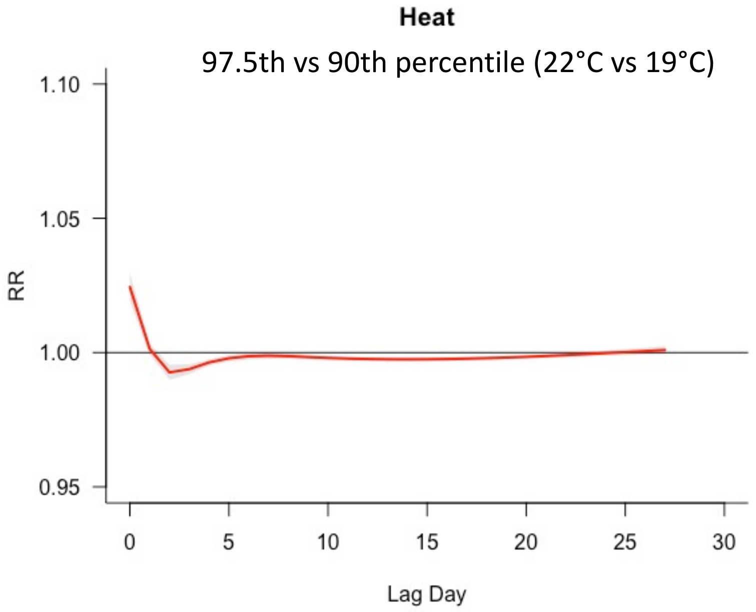


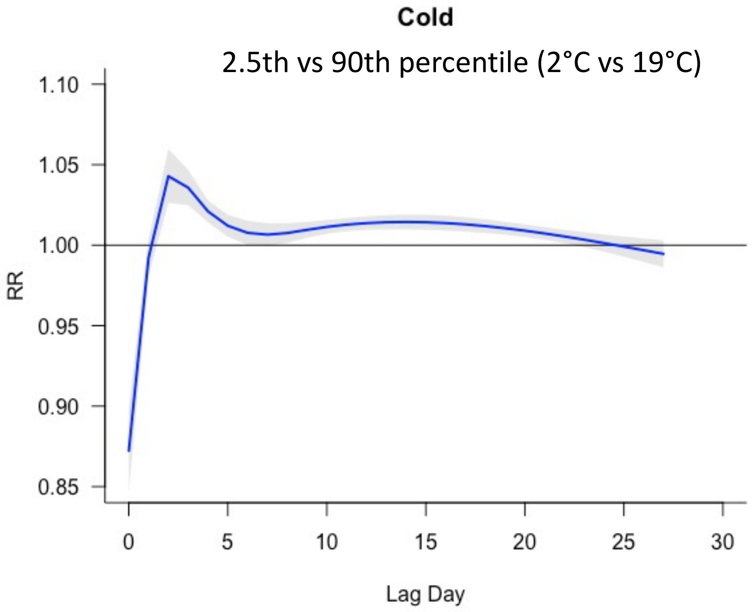
Summer heat and daily mortality, London 2002-2006 (June – September).
Temperature is linear, adjusted for ns(day-of-season, df=4)*year, and year.

Distributed lag structure

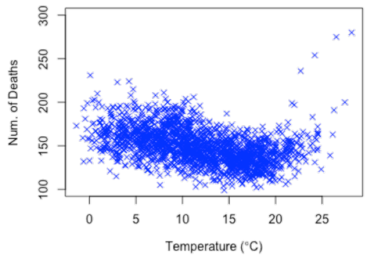
Temperature and mortality, London 2002-2006 (full year data)

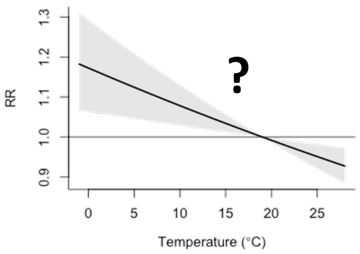
- Hot and cold temperatures





- But distributed lag model ***assume linear*** exposure-response relationship
- Linear temperature-mortality association is a strong assumption!
- Need more flexible method than the above



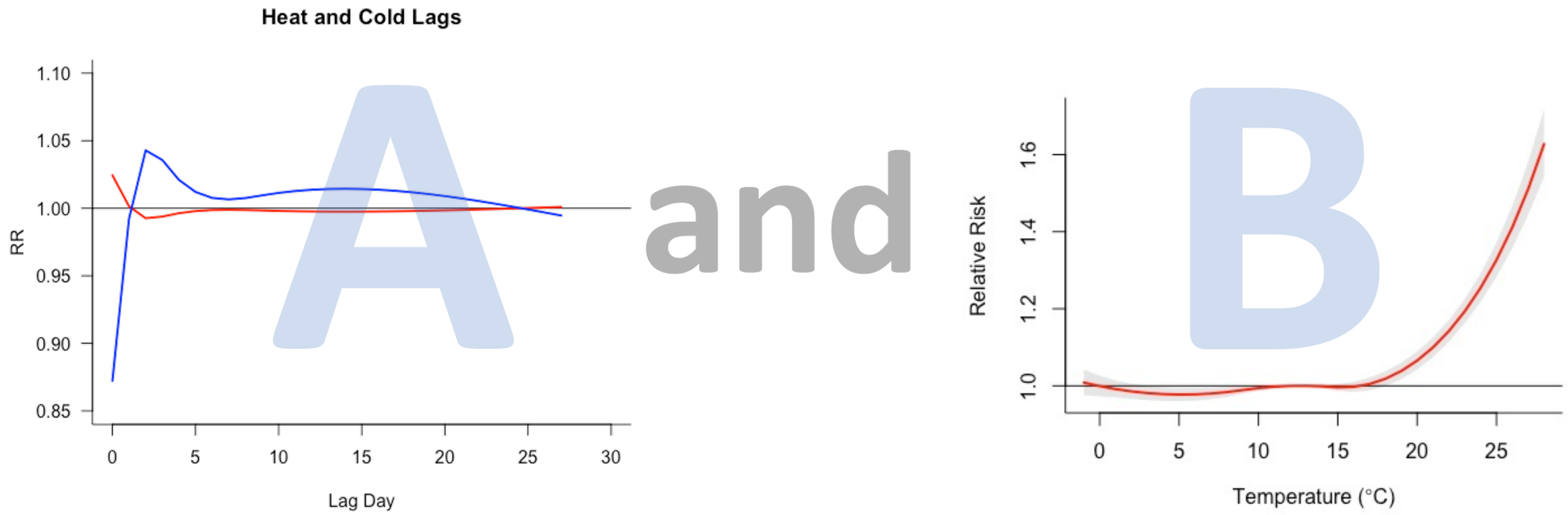


Distributed lags for non-linear association

A. **Lag-response** association: Distributed Lag Model

B. **Exposure-response** association can be linear or non-linear

Need a flexible method to combine the two dimensions, **A** and **B**



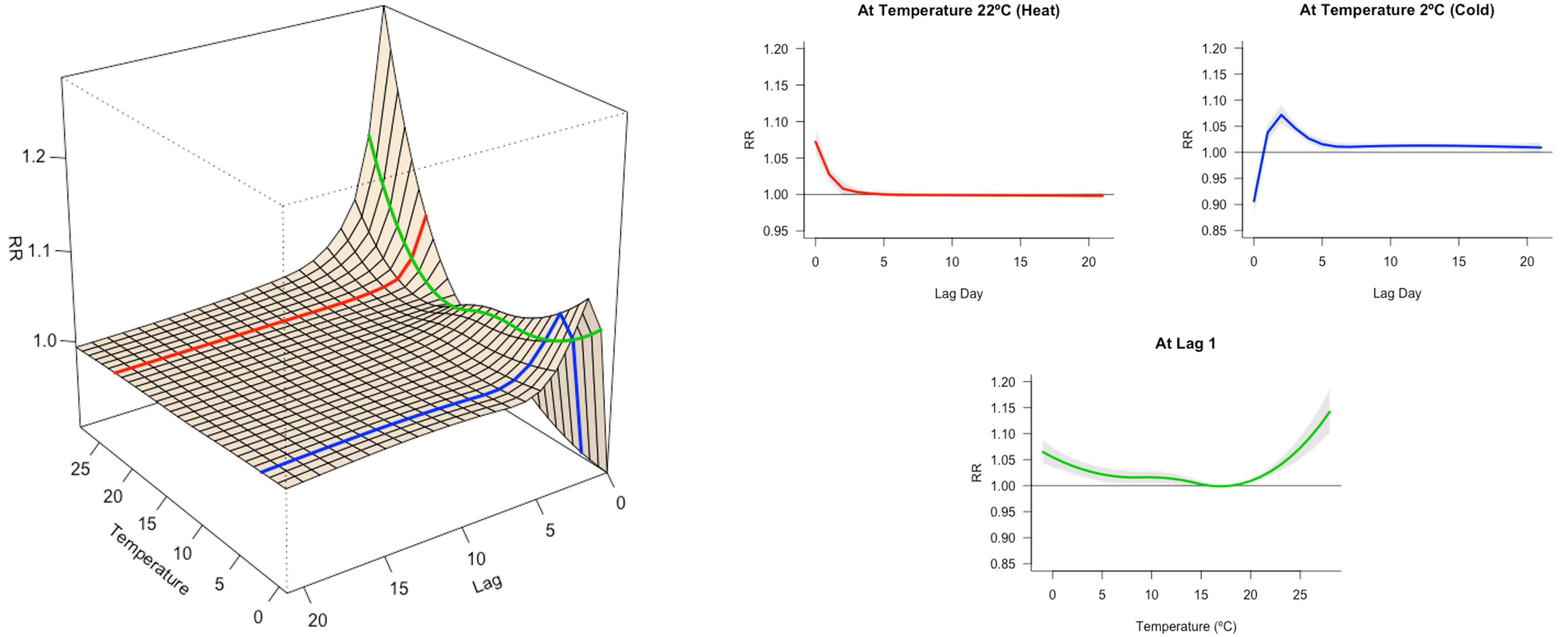
- Distributed lag non-linear model (DLNM)

Distributed lag nonlinear model (DLNM)

- DNLM is represented by a bi-dimensional function, consisting of exposure-response function ***g(x)*** and lag-response function ***w(l)***

$$\log[E(Y_t)] = \alpha + \sum_{l=l_0}^{L} g \cdot w(x_{t-l}, l, \theta) + cov$$

Temperature and mortality, London 2002-2006



3-D representation of exposure-lag-response risk surface

Overall cumulative exposure-response curve across all 21 lags

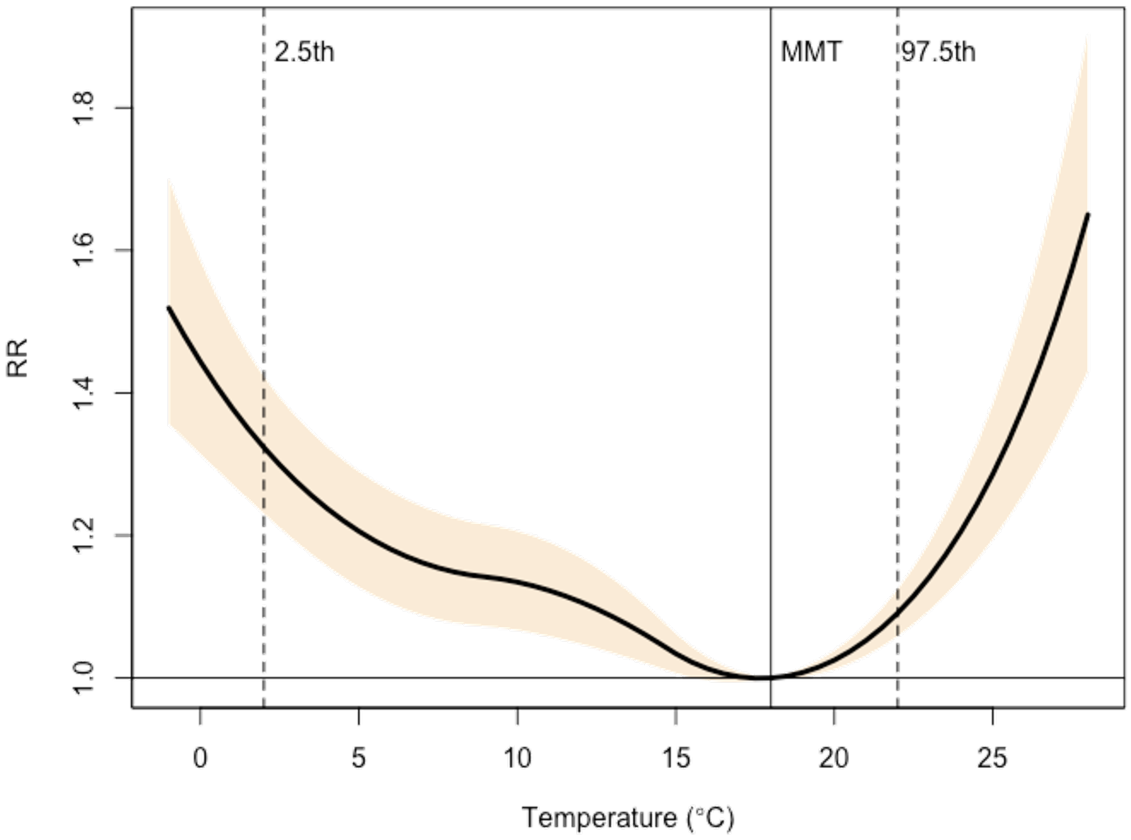
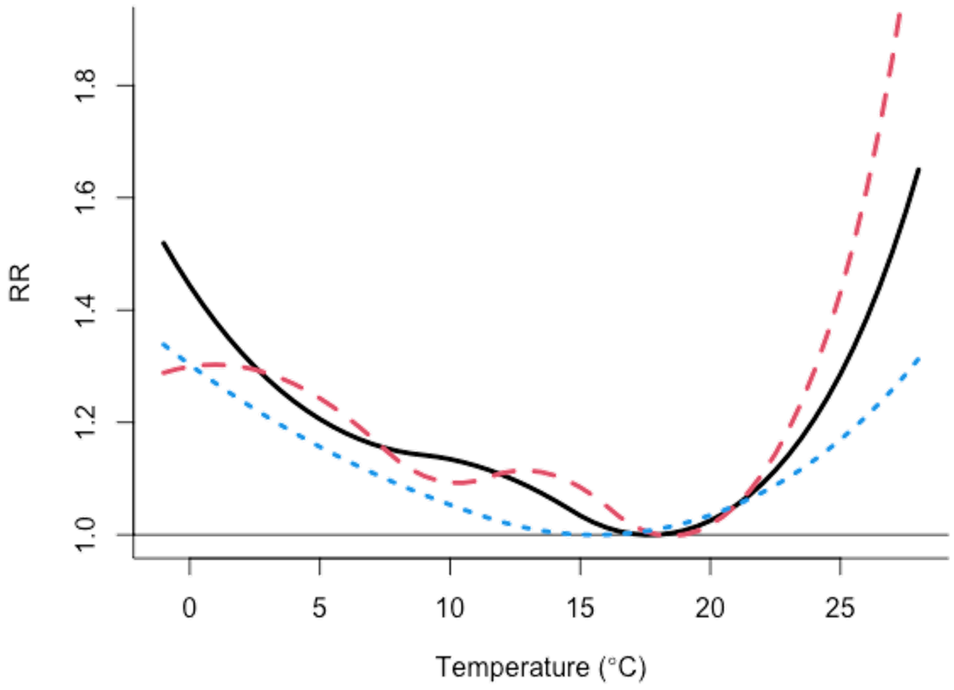


Table. Overall cumulative relative risk (RR) and the 95% confidence intervals (CI) for heat and cold

Exposure	RR	95% CI
Heat	1.09	(1.06, 1.12)
Cold	1.32	(1.23, 1.42)

Cold at 2°C (2.5th percentile)
Heat at 22°C (97.5th percentile)
Reference MMT at 18°C (85th percentile)

Sensitivity: number of knots for exposure-response curve

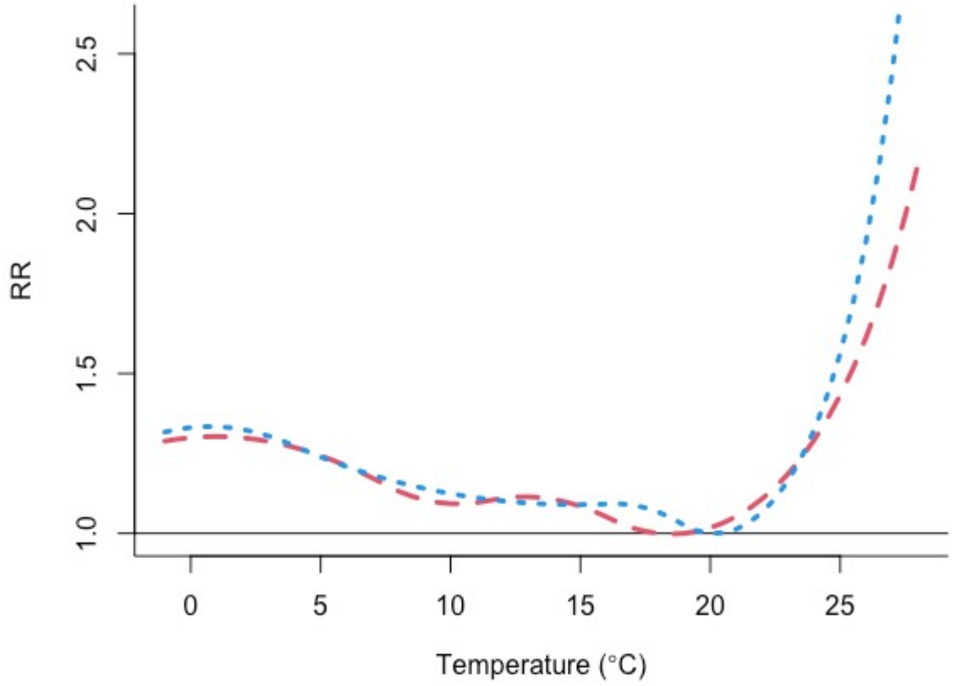


No. of knots	Location (percentile)	qAIC	Sum of absolute PACF	MMT (°C)
3	25, 50, 75	1937	0.637	18
2	33.3, 66.7	1958	0.634	18
1	50	2010	0.706	16

	RR (95% CI)	
No. of knots	Heat	Cold
3	1.11 (1.07, 1.14)	1.30 (1.21, 1.40)
2	1.09 (1.06, 1.12)	1.32 (1.23, 1.42)
1	1.08 (1.03, 1.12)	1.24 (1.17, 1.31)

Q: In table above, why are the RRs for heat smaller? The plot seems to indicate larger values.

Sensitivity: position of knots for exposure-response curve

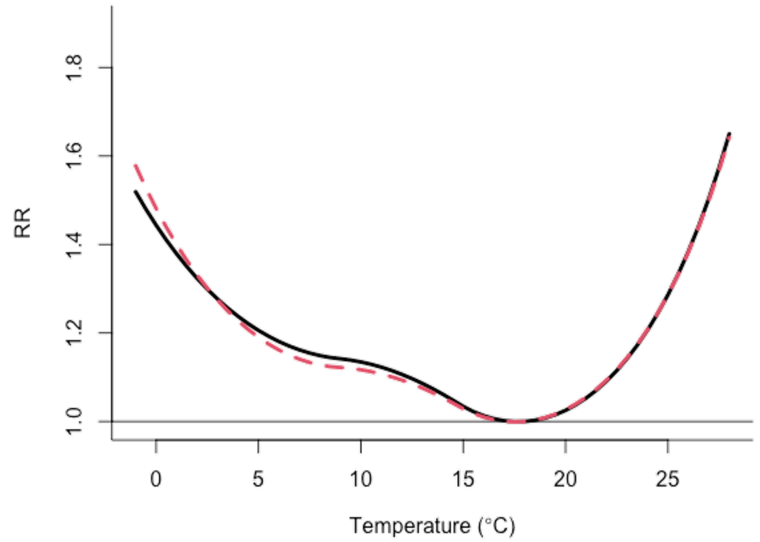


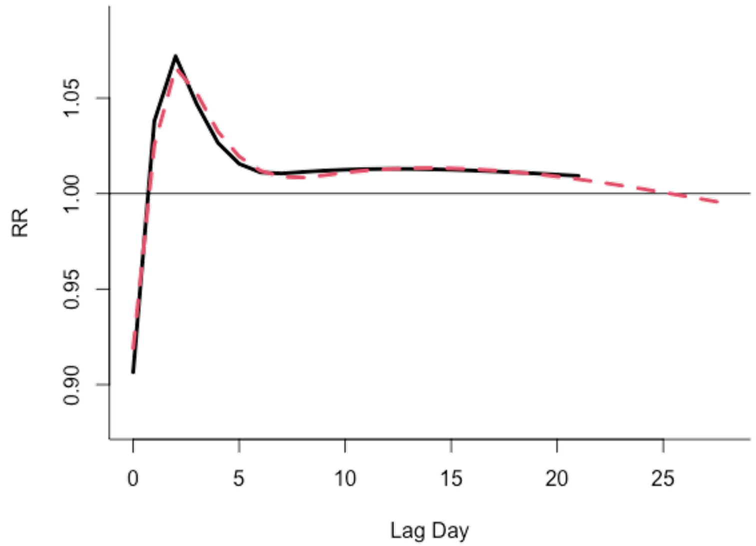
No. of knots	Location (percentile)	qAIC	Sum of absolute PACF	MMT (°C)
3	25, 50, 75	1937	0.637	18
3	10, 75, 90	1931	0.654	20

		RR (95% CI)	
No. of knots	Location of knots	Heat	Cold
3	25, 50, 75	1.11 (1.07, 1.14)	1.30 (1.21, 1.40)
3	10, 75, 90	1.07 (1.05, 1.09)	1.32 (1.23, 1.43)

Q: Consider a more extreme definition for heat. (99th vs 90th percentile). Will the results be more sensitive to placement of knots?

Sensitivity: duration of lag period

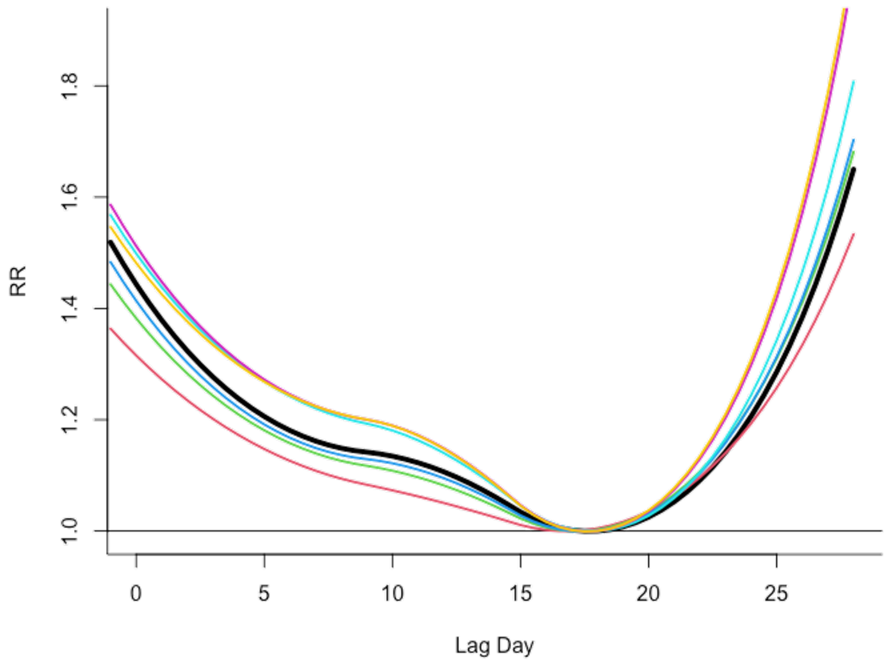




Lag period (days)	qAIC	Sum of absolute PACF	MMT (°C)
21	1958	0.634	18
28	1953	0.637	18

Q: What about 14 lag days?

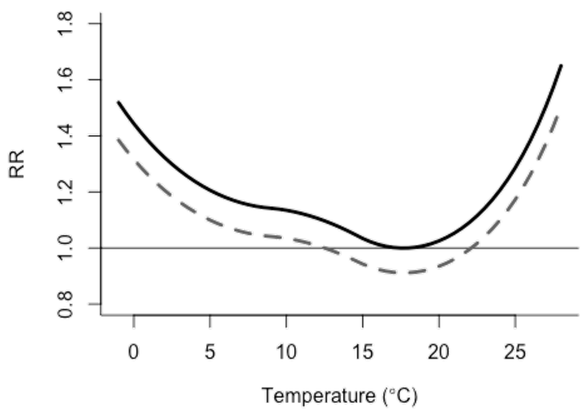
Sensitivity: degree of smoothness for time adjustment

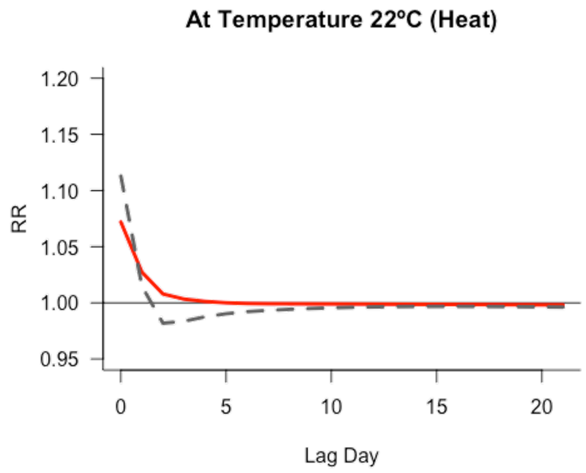


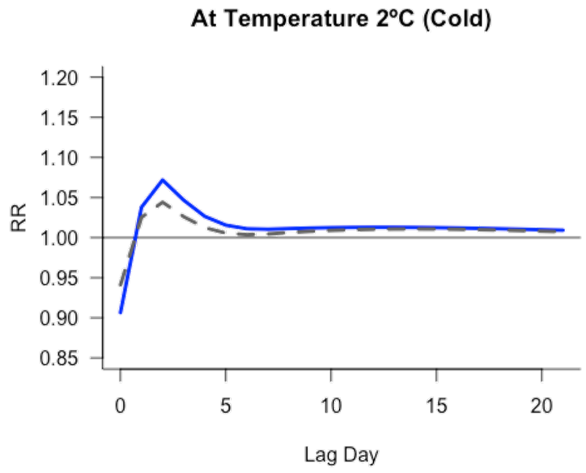
DF for time*	qAIC	Σ|PACF|	MMT (°C)	RR (95% CI) Heat	RR (95% CI) Cold
3	2014	0.797	17	1.10 (1.07, 1.13)	1.23 (1.16, 1.32)
4	1967	0.658	17	1.11 (1.07, 1.14)	1.28 (1.21, 1.37)
5	1962	0.628	17	1.10 (1.07, 1.14)	1.30 (1.22, 1.39)
6	1958	0.634	18	1.11 (1.07, 1.14)	1.30 (1.21, 1.40)
7	1950	0.714	18	1.11 (1.07, 1.14)	1.39 (1.29, 1.49)
8	1944	0.840	18	1.13 (1.09, 1.17)	1.39 (1.28, 1.51)
9	1938	0.980	18	1.13 (1.09, 1.18)	1.38 (1.27, 1.50)

* Degree of freedom (DF) for natural cubic spline of time per year

Choosing a centering value as the reference







- Choose a suitable reference based on understanding of exposure and its relationship with outcome
- Above figures for temperature: dashed lines centered at 12.5 °C, while solid lines at 18°C

Q: What is a good reference for air pollutant?

- Different centering → different results → different interpretation

Summary

- Distributed lag models for lag-response association
- Extension to include nonlinear exposure-response association
  - bi-dimensional exposure-lag-association using DLNM
  - Example using London data
- Common issues and suggestions for sensitivity checking

References

1. Armstrong B. Models for the relationship between ambient temperature and daily mortality. *Epidemiology* 2006;17(6): 624-631.
2. Gasparrini A, Armstrong B, and Kenward MG. Distributed lag non-linear models. *Statistics in Medicine* 2010;29(21): 2224-2234.
3. Gasparrini A. Modeling exposure-lag-response associations with distributed lag non-linear modelsl. *Statistics in Medicine* 2013;33(5):881-99.
4. Gasparrini A. Distributed Lag Linear and Non-Linear Models in R: The Package dlnm. *Journal of Statistical Software* 2011;43(8): 1-20.
5. Schwartz J. The distributed lag between air pollution and daily deaths. *Epidemiology* 2000;11(3):320-326.

\-