Modelling of complex, non-linear relationships in time series data while accounting for delayed effects 2

Robbie M Parks, PhD

22nd July 2025

Email: robbie.parks@columbia.edu

BlueSky: @robbiemparks

Website: sparklabnyc.github.io

Art by Amy Wolfe

COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

Outline from previous lecture

- Non-linear exposure-response curves
- Linear regression as an assumption
- Polynomials
- Splines
- Piecewise linear splines
- Natural splines
- Penalized splines
- Which to use?

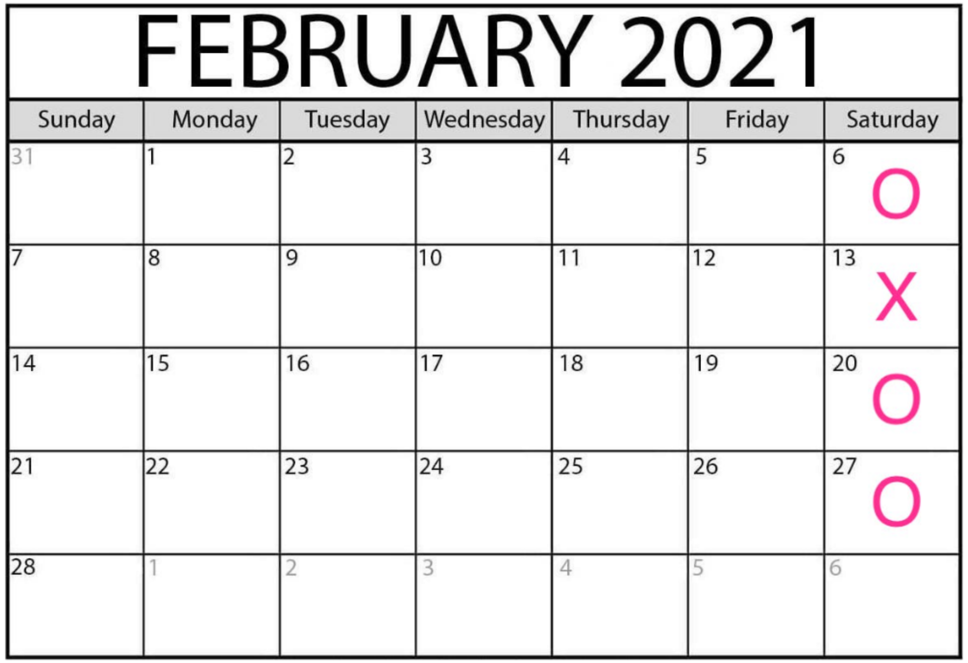
Outline

- Case crossover design
- Time series design

Time-stratified case-crossover design

1. What does it do?
   - "A method for studying transient effects on the risk of acute events" (Maclure, 1991)
   - Compares a case's exposure during case-defining event with that same person's exposure at otherwise similar "reference" times

2. Why is this useful?
   - Only examines cases
   - Each person acts as own control
   - No confounding by time invariant variables
   - Saves effort and time



Time-stratified case-crossover design

1. ”There can be no confounding by time invariant variables” (Maclure, 1991)
   - What is time-invariant?
     - Sex?
     - Socio-economic status?
     - Education?
     - BMI?
     - Smoking?
     - Other lifestyle factors?

- Above therefore cannot be confounders by design if reasonably thought of as time invariant
  - Time scale important!

Time-stratified case-crossover design

1. So what *still* could be a confounder?
   - Time-varying variables
     - Day of week, season, long-term trends?
     - Other exposure variables

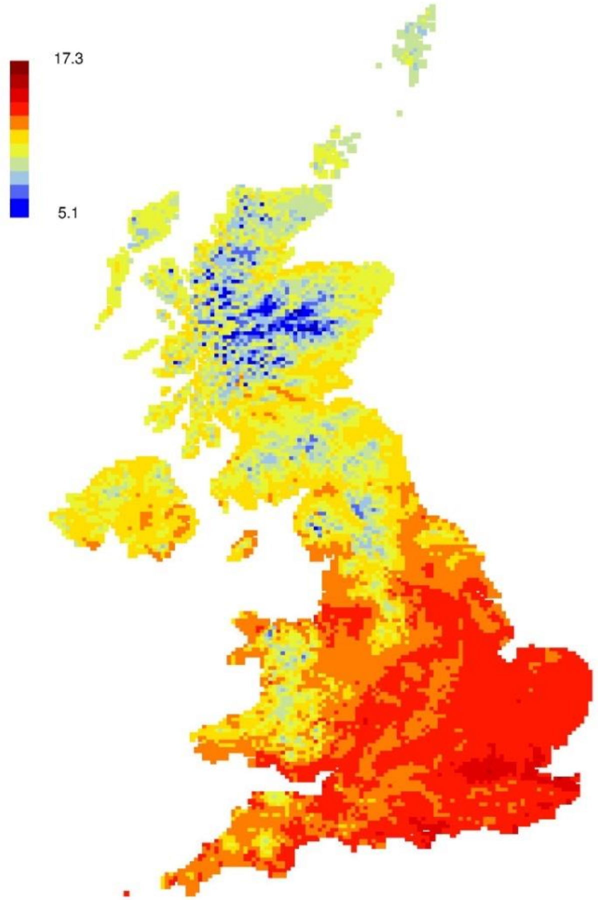
2. How to deal?
   1. Adjust as usual in model (i.e., additional terms)
   2. Also can further match by exposures
      - But lose control days

      \+ Good confounding control

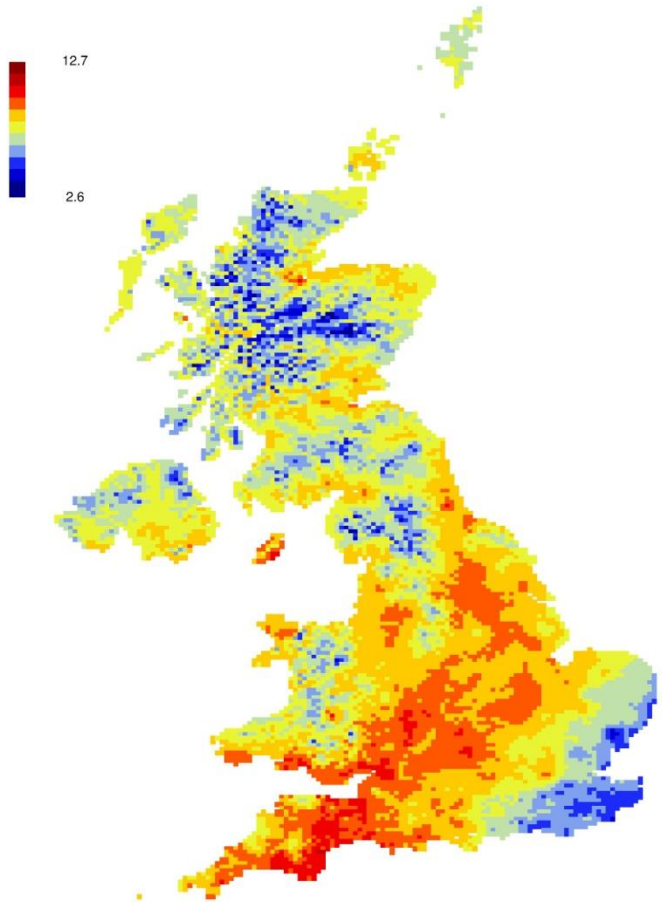
      \- Loss of power to make inferences

Time-stratified case-crossover design

Mean daily temperature, °C
May-Sept, 2001-2004



Mean daily temperature, °C
12 May 2004

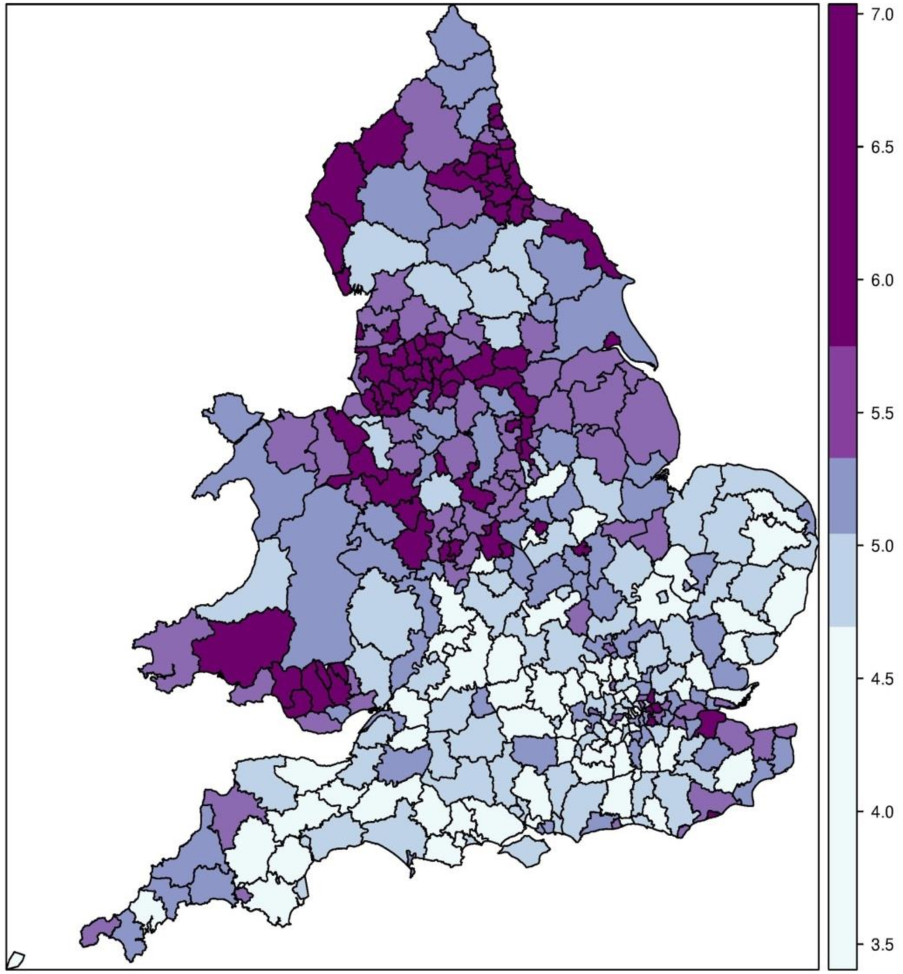


Bennett et al., *Nature Climate Change*, 2014

Time-stratified case-crossover design

- Cardiorespiratory deaths
  - (ICD 10 = I*,J*)
  - Total deaths 2001-2004 = 406,697
- Age, sex at individual level
- Mean daily temperature at 5km grid linked via postcode
- Confounders: Pollution (PM_{10}), national holidays

Average cardiorespiratory death rates 2001-2004



Bennett et al., *Nature Climate Change*, 2014

Time-stratified case-crossover design

- Case crossover design
  - Each case serves as its own control
  - Compare the temp on the day of death with the temp on other days “near” to the day of death
  - Pick control dates on same day of week, within same month of death, either side of Date of death e.g.

    |---C------D------C------C-----|

    |-C------C------D------C------C|

- Separate analyses for hot and cold months
- Separate analyses by sex and

  age group <=74, 75-84, >=85

- Estimate the percentage increase in odds of death per degree increase in temp

Bennett et al., *Nature Climate Change*, 2014

Time-stratified case-crossover design

- Case crossover design analysed using conditional logistic regression:

$Y_{\text{indiv i,case/control j}} \sim \text{Pois}(\lambda_{ij})$

$\log(\lambda_{ij}) = f(\beta_{0\,d}, \beta_{1\,d}, \text{Temp}_{ij}) + \text{Confounders}_{ij} + \delta_i$

δ_i = parameter linking the case/controls of indiv. i

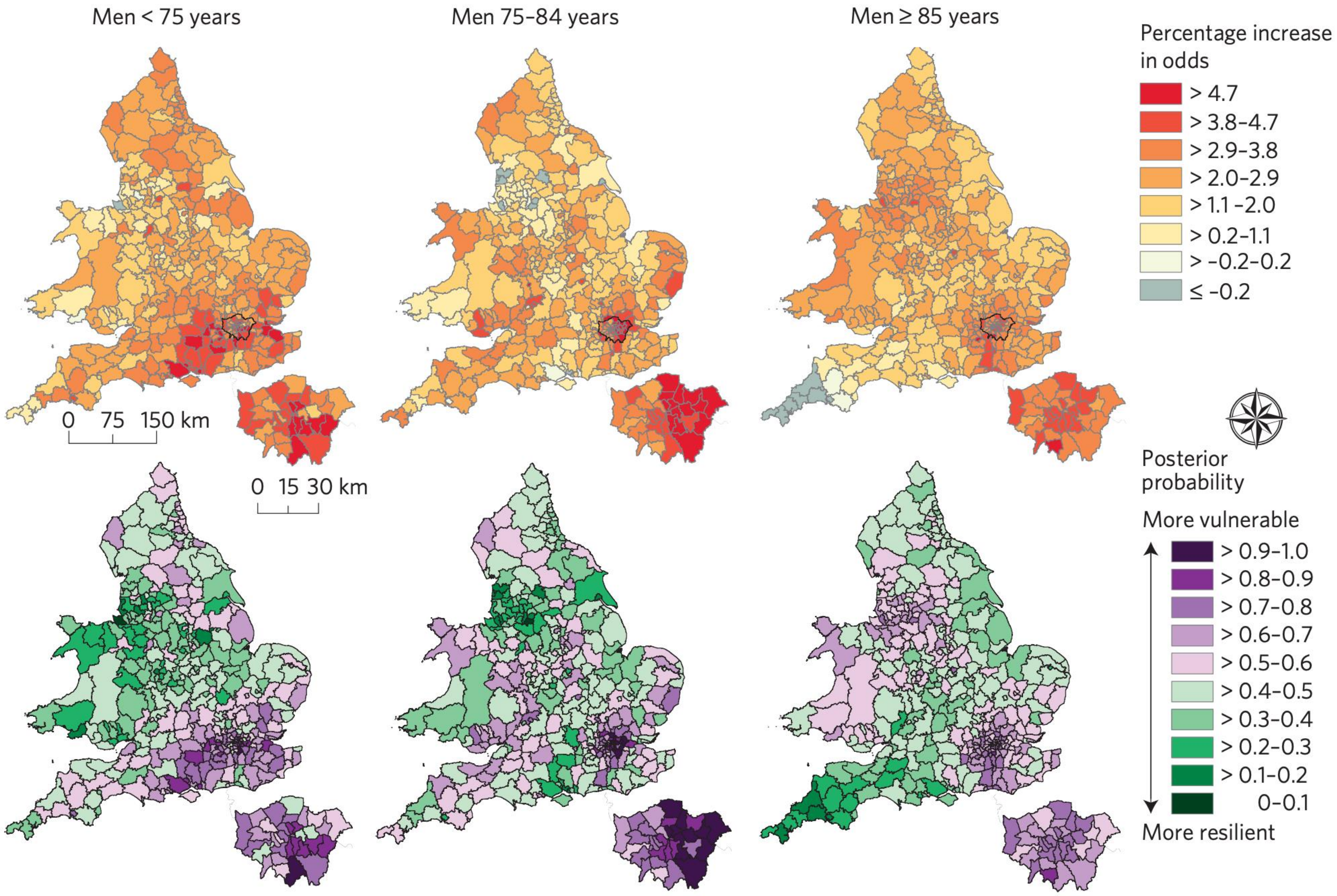
$\beta_{0\,d}$ = Fixed Threshold for cases in District d

$\beta_{1\,d}$ = Slope (supra-threshold) for cases in District d

= Structured + Unstructured District RE's (BYM)

Bennett et al., *Nature Climate Change*, 2014

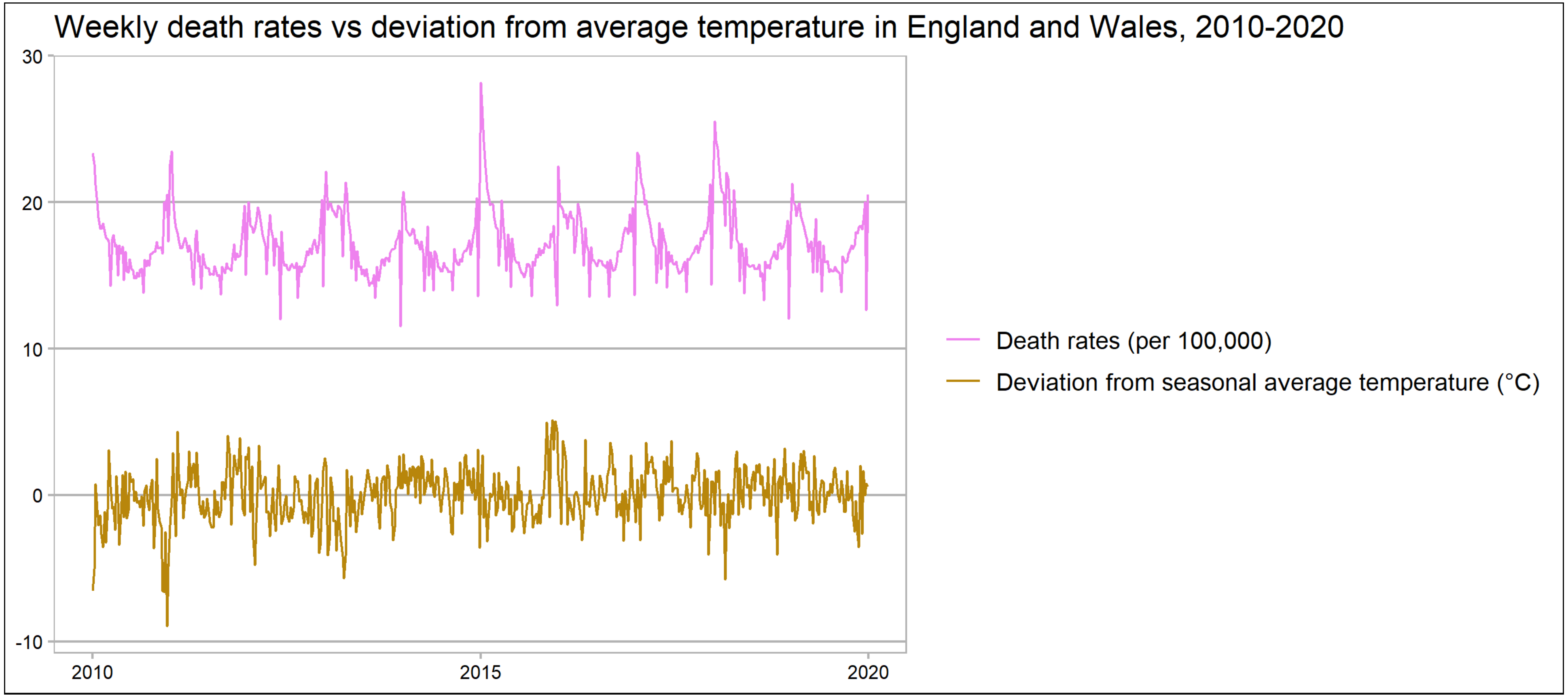
Time-stratified case-crossover design



Bennett et al., *Nature Climate Change*, 2014

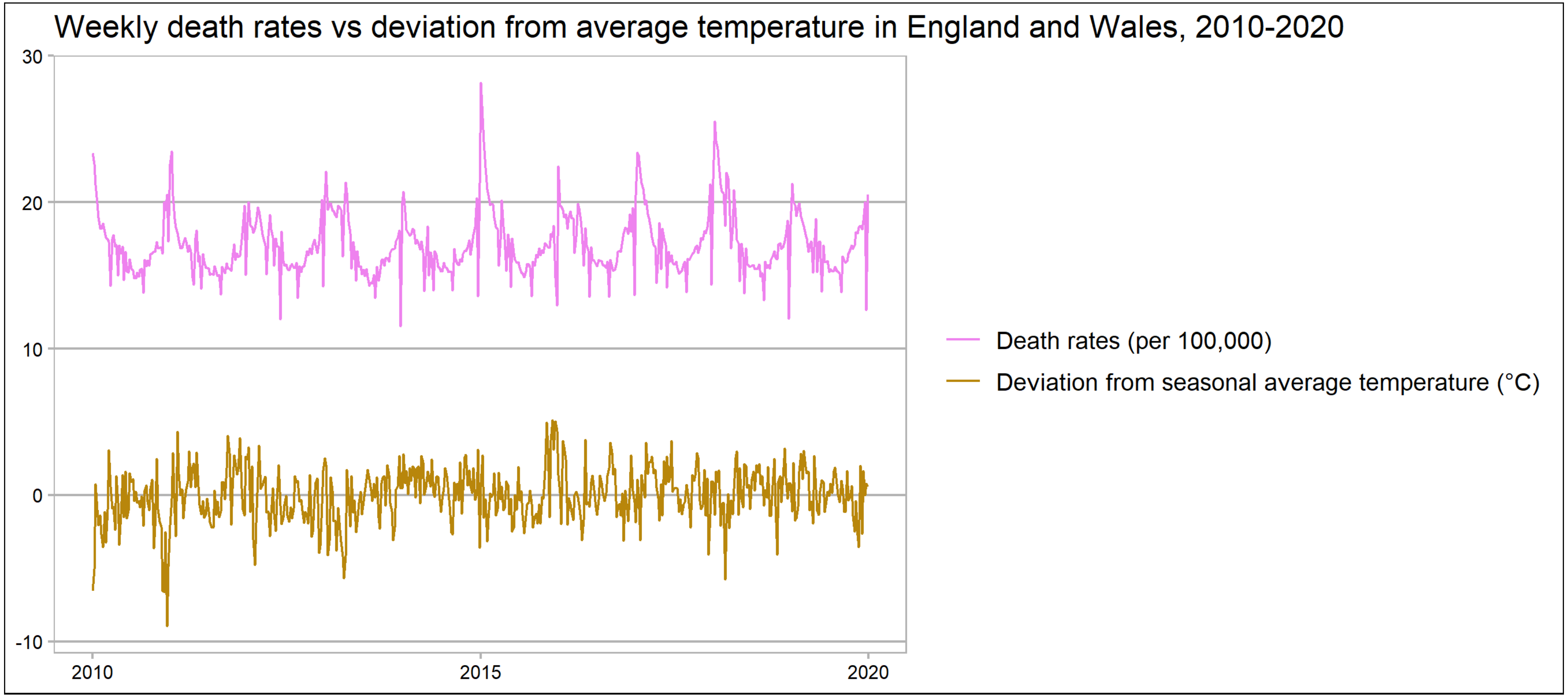
Time series design

1. What does it do?
    - Examines how variation in exposure over time is associated with variation in number of health outcome events over time



Time series design

2. How is this different from case-crossover design?
   - Unit of analysis is time unit rather than case
     - Implications for confounding association of interest



Time series design

- How does it work?
  1. Obtain counts of health outcomes of interest in time unit in each geographic unit (if spatial)
     - For example, deaths by year in each county in the United States (1999-2015)
  2. Obtain concentrations of exposure of interest in each geographic unit (if spatial)
     - For example, annual concentrations of $PM_{2.5}$ by county in the United States (1999-2015)
  3. Adjust by appropriate covariates
  4. Run Poisson model (if counts) with log-link function

Time series design

- In time series design, what could be a confounder?
  1. Individual-level factors?
     - No, as now unit is not individual, because time the unit of analysis, not individual
  2. Population characteristics by spatial unit?
     - Yes, if you can argue that they change by the unit of analysis (e.g., year)
       - E.g., does %poverty in a US county change appreciably year-to-year?
  3. Characteristics of spatial unit itself?
     - Potentially, if they change by unit of analysis
       - E.g., does %urban in a county change appreciably year-to-year?
  4. Temporal trends
     - Long-term trends
     - Seasonality
  5. Other exposures
     - If pollution is exposure of interest, do we also have to include temperature?

Time series design

- Outcome: deaths by underlying cause of death from vital registration and population from census

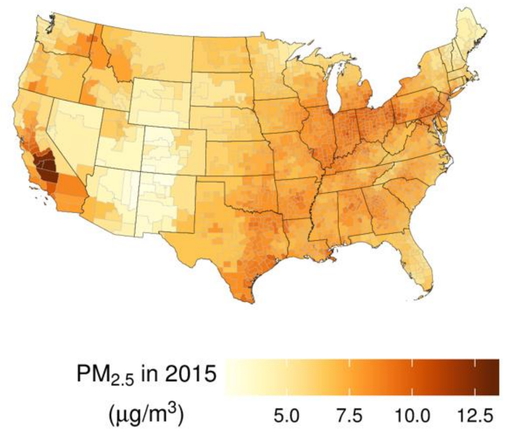
- Exposure: annual mean $PM_{2.5}$ (LUR model assimilating observations in universal kriging framework)

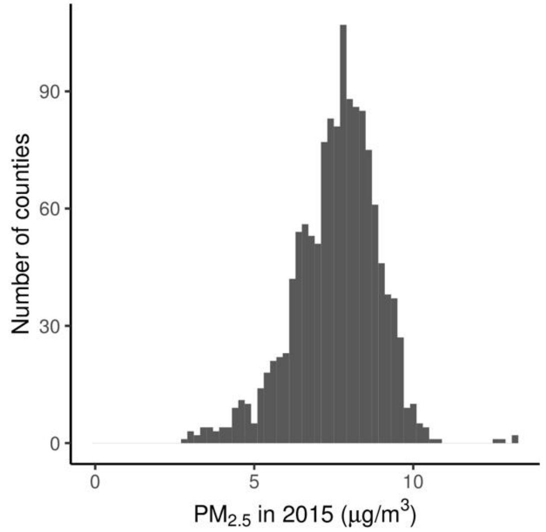
- Covariates: % poverty, % black, % high-school graduates,
  % living in urban areas, % unemployed, per capita income,
  age-standardised lung-cancer death rates as proxy for smoking,
  annual mean temperature, annual mean relative humidity

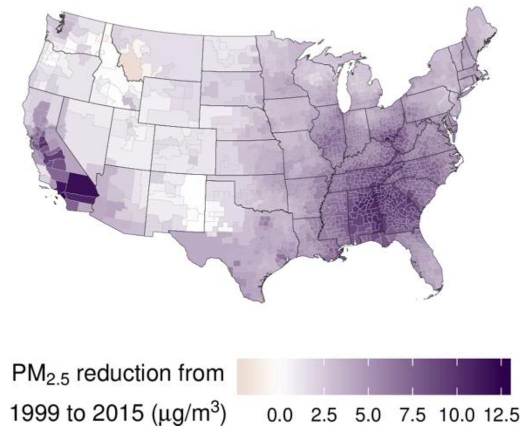
Bennett et al., *PLOS Medicine*, 2019

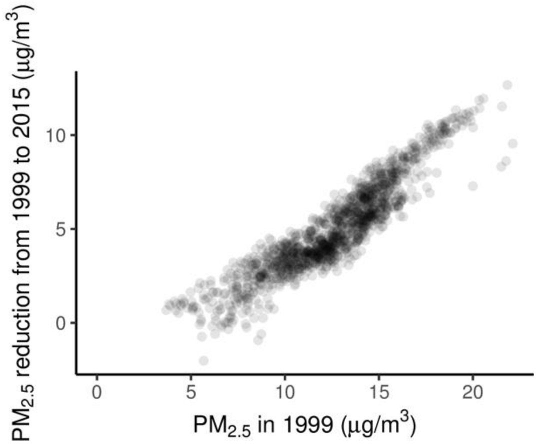
Study population, outcome and exposure assessment

- Study population
  - 18.4 million cardiorespiratory deaths (1999-2015)
  - Entire contiguous United States
- Outcome assessment
  - Annual county-level, cardiorespiratory death counts by principle cause of death diagnosis (I/J ICD-10)
- Exposure assessment:
  - Annual mean $PM_{2.5}$ (1999-2015)
  - LUR model assimilating observations in universal kriging framework









Bennett et al., *PLOS Medicine*, 2019

Statistical model

- Association between annual cardiorespiratory death rates and $PM_{2.5}$
  - Bayesian spatio-temporal model with Poisson counts
  - Leverages variations over space and time to infer associations

- To quantify association:
  - Death rate ratio per $10\mu g/m^3$

- Adjusted for:
  - Longer-term time trends
  - County specific effects

Bennett et al., *PLOS Medicine*, 2019

Time series design

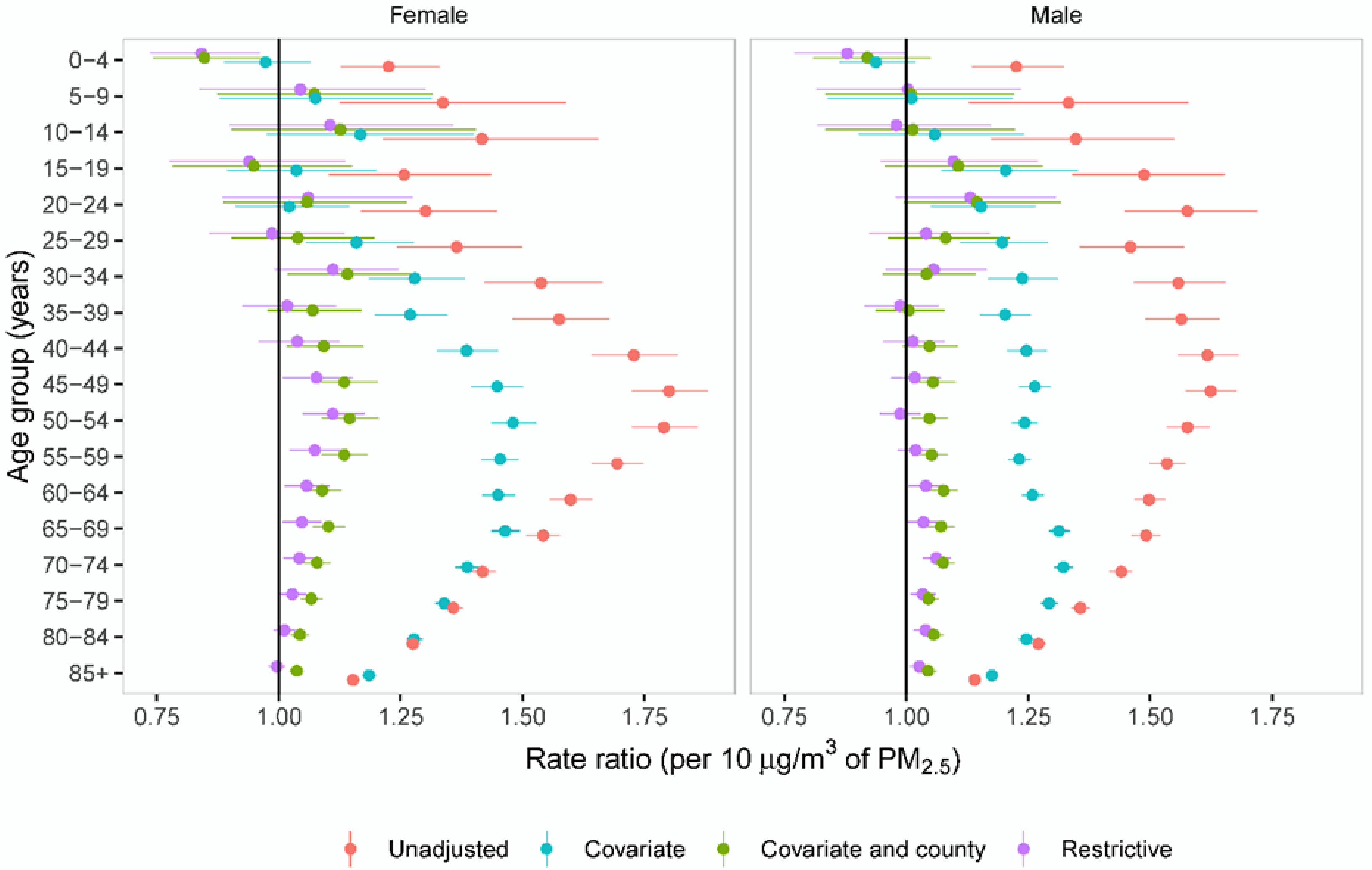
For each age-sex group and where *c = county, t = year*

$$Deaths_{tc} \sim Poisson[population_{tc}.\,deathrate_{tc}]$$

$$\begin{aligned}
\log(d_{tc}) \quad = \quad & (\alpha_0 + \beta_0.\mathrm{t}) && \text{common terms} \\
& + \nu_t && \text{non-linear time term} \\
& + \boldsymbol{\gamma}.\,PM_{tc} && \text{pollution term} \\
& + \textstyle\sum_{i=1}^{i=9} \theta_i \cdot X_{itc} && \text{covariates} \\
& + \alpha_c && \text{county random effect} \\
& + \epsilon_{tc} && \text{over dispersion}
\end{aligned}$$

Bennett et al., *PLOS Medicine*, 2019

Effect size for cardiorespiratory deaths



Bennett et al., *PLOS Medicine*, 2019

Statistical model

- Association between monthly injury death rates and anomalous temperature
  - Bayesian spatio-temporal model with Poisson counts
  - Leverages variations over space and time to make inferences

- To quantify association:
  - Age-group-sex-month-cause-specific death rate ratio per 2°C warm anomaly

- Adjusted for:
  - Longer-term time trends
  - State and month specific effects

Parks et al., *Nature Medicine*, 2020

Time series design

For each age-sex group and where *s = state, t = time*

$$Deaths_{ts} \sim Poisson[population_{ts}.\, deathrate_{ts}]$$

$$\begin{aligned}
\log(d_{ts}) \quad = \quad & (\alpha_0 + \beta_0.\mathrm{t}) && \text{common terms} \\
& + (\alpha_s + \beta_s.\mathrm{t}) && \text{state random effects} \\
& + (\alpha_m + \beta_m.\mathrm{t}) && \text{month random effects} \\
& + \zeta_{\mathrm{sm}} + \psi_{\mathrm{sm}}.\mathrm{t} && \text{state-month interactions} \\
& + \nu_t && \text{non-linear time term} \\
& + \boldsymbol{\gamma}.\, Anomaly_{ts} && \text{Temperature anomaly terms} \\
& + \epsilon_{tc} && \text{over dispersion}
\end{aligned}$$

Parks et al., *Nature Medicine*, 2020

Comparing case-crossover (CC) with time series (TS)

- CC can use individual level exposure data
- TS requires spatial unit-aggregate exposure
- CC can assess effect modification by individual-level data
  - E.g., smoking, BMI
- TS can do this too but careful about interpretation
- CC arguably easier to control for confounding by design (matching)
- TS great care required when building model

Comparing case-crossover (CC) with time series (TS)

- When could CC and TS given (nearly) same results?
- Equivalent if:
  - Spatial unit averages from TS used as exposures in CC (i.e., not individual)
  - Models both correctly specified
  - All things being equal, CC would give wider CIs because of less power

Outline

- Case crossover design
- Time series design

Getting ready for the lab

- This lab will involve taking some models and concepts from the **Modelling of complex, non-linear relationships in time series data while accounting for delayed effects 2** lecture and introduce you to the way case crossover and time series design works.

Application

- How can you imagine applying this learning to your data and your research questions?

Questions

- Questions?

Art by Amy Wolfe

Modelling of complex, non-linear relationships in time series data while accounting for delayed effects 2

Robbie M Parks, PhD

22nd July 2025

Email: robbie.parks@columbia.edu

BlueSky: @robbiemparks

Website: sparklabnyc.github.io

COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH