# Practical day 2 – Modelling Associations Between Climatic Factors & Health Outcomes Using Aggregated-Level

Note: the practical consists of 19 questions in this word document, to answer them you will code using the R script. The code is already written but the questions will mainly require reasoning.

The data you will be using is open-source data coming from Eurostat (weekly all-sex, all-age, all-cause mortality counts at NUTS3 level) and Copernicus Climate Change Service (hourly gridded 2-m temperature data from the high-resolution ERA5-Land reanalysis into weekly regional averages of daily mean 2-m temperature). This data has been processed by the Adaptation team at ISGlobal (Barcelona), funded by the European Research Council Consolidator Grant EARLY-ADAPT No 865564, <https://www.early-adapt.eu/>. The data provided is for the countries of Sweden, Spain, and France from January 2000 to December 2019.

The dataset contains the following columns:

* date – The calendar date of observation.
* mort – Weekly count of all-cause mortality for the region.
* temp – Mean weekly air temperature (°C) for the region.
* wop – Week-of-period variable, used to model long-term and seasonal trends with splines.
* location – Regional identifier matching each location in the analysis (e.g., city, province).

In this practical you will:

1. Prepare and structure data for multiple regions, including mortality counts, temperature, and seasonality variables.
2. Fit generalized linear models (GLMs) with cross-basis functions to model the non-linear and lagged association between temperature and mortality.
3. Identify the Minimum Mortality Temperature (MMT) and center the exposure–response curves.
4. Extract reduced coefficients and covariance matrices for each region to use in a multivariate meta-analysis.
5. Visualize results, including: Time series of observed vs. seasonal mortality, Regional cumulative exposure–response curves before and after pooling.
6. Export results for interpretation, model selection and reporting.

**Data exploration and preparation**

Understanding the dataset is a crucial first step before fitting models. This exploration helps to check the structure, identify potential issues and summarize key variables like mortality, temperature, and time periods.

1. How many rows and columns are in the dataset?
2. How many regions are included? (in total and per country?)
3. What is the total mortality across all regions? And per country?
4. What is the time range (earliest and latest dates) covered by the data (total and per country)?
5. What does the distribution of temperature (temp) look like overall and by country?
6. Plot the time series of weekly mortality and of weekly mean temperature for an average year. Do you notice any seasonal patterns in either series?

**First-Stage Modelling**

In this step, we fit region-specific models to estimate the association between weekly mean temperature and mortality, accounting for both non-linear effects of temperature and delayed (lagged) effects over several days. This stage produces region-level estimates (coefficients and covariance matrices) that will later be combined in a multivariate meta-analysis.

1. What type of regression model is used in this stage, and why is it suitable for mortality counts?
2. Why do we include a seasonality term (ns(wop, df = ...)) in the model?
3. What are the two dimensions that are combined in a cross-basis function?
4. Why is it important to include lagged effects when modelling temperature and mortality?
5. How are the location of knots for temperature defined in the script?
6. What do the cen and at arguments control?
7. What’s the purpose of centering at the Minimum Mortality Temperature (MMT)?
8. How does the model determine the Minimum Mortality Temperature (MMT)?
9. Why do we need to save both the coefficients and covariance matrices for the second-stage meta-analysis?

**Second-Stage: Meta-Analysis and Interpretation**

In the second stage, we combine the region-specific first-stage estimates using multivariate meta-analysis. This step allows us to obtain pooled exposure-response curves, assess heterogeneity between regions, and explore the impact of meta-predictors (e.g., temperature variability, socio-economic indicators).

1. Based on the Wald-test results, do all meta predictors explain part of the association? Assess their significance.
2. What differences do you observe between the regional (non-meta) and meta-analysis plots?
3. Compare regions of Sweden, France, and Spain. What differences appear in the exposure-response curves? How does the regional cumulative curve differ?
4. Interpret the final cumulative curve for the three countries.