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

Note: the practical consists of 18 questions in this word doc, to answer them you will code using the r script. Some code is filled out, other code still needs to be written. The answer key is provided as an R Markdown file.

We have simulated primary care and hospitalisations data of people with asthma exacerbations in a district of Sweden from January until December 2018, that was linked to environmental data for a case time series format in the previous practical on Linking Climate Data with Health Data for DLNM Analysis**.**

The dataset contains the following columns:

* id – A unique identifier for each individual in the study. Each person contributes multiple rows of follow-up data.
* date – The calendar date of observation for that individual.
* temp – The mean daily air temperature (°C) for that individual’s area on the given date.
* smoking – Indicator of whether the individual is a smoker (1 = smoker, 0 = non-smoker). This does not vary across time for the same person.
* outcome – Indicator of whether an asthma exacerbation occurred on that day (1 = yes, 0 = no).

In this practical you will 1. Explore an individual-level case time series dataset of asthma exacerbations, 2. Learn how to model the association between daily mean temperature

and asthma exacerbations using cross-basis functions, 3. Study subgroup differences (e.g., smokers vs non-smokers), 4. Predict and interpret exposure–response curves.

**Data exploration**

Exploring the case time series dataset is an important first step before doing any formal analysis. It helps us understand what information is available, how it is structured, and what the data look like in practice. We summarise the number of individuals, follow-up time, outcomes, and exposures such as temperature or smoking, to get a clearer picture of the study setting.

1. How many rows and columns are in the dataset?
2. How many unique individuals are there?
3. What does the distribution of temperature (tmean) look like?
4. How many days of follow-up are there on average per individual?
5. On average, how many asthma exacerbations does each person have?
6. How many individuals had no outcomes at all? Remove them from the dataset.
7. How many individuals are smokers (ever smoking == 1)?
8. Do smokers have more exacerbations on average than non-smokers?
9. Explore the data for one individual:
   1. Plot the time series of asthma exacerbations.
   2. Plot the time series of mean daily temperature.
   3. Is the number of asthma exacerbations realistic?
   4. Does there seem to be a correlation between temperature and asthma exacerbations?

**Data modelling**

In the next step we want to model the association between daily mean air temperature and asthma exacerbations. To do this we use a case time series approach. This type of model is well suited for individual-level longitudinal data because it allows us to compare case days with control days within the same person, while accounting for both time and lagged effects of exposure.

1. What are knots, and why do we need them when modelling temperature using splines?
2. How do we define the knots and centering values for temperature?
3. What is the stratum variable, and why do we need it?
4. Which two dimensions are combined in a cross basis?
5. Create the crossbasis using the code in the script, for 21 lag days. What arguments do we hand crossbasis()?
6. Fit a Poisson regression model with the crossbasis as exposure, day of week as a covariate, and the stratum variable as fixed effects.

**Crossprediction**

Once the model is fitted, we use a crossprediction to translate the estimates into an interpretable curve. The crossprediction shows the relative risk of asthma exacerbations across the observed range of daily mean temperatures, compared to a chosen reference (the centering value, in our case the median). This allows us to visualise and interpret the exposure–response relationship, and later compare it between subgroups such as smokers and non-smokers.

1. Predict and plot the exposure–response curve for temperature from our model using the provided code.
2. How do we interpret the exposure–response curve?
3. Compare the temperature–asthma exacerbation association between smokers and non-smokers by running the provided code. Interpret the plot.