Burden of disease due to environmental exposure

An analysis of attributable risk due to air pollution and temperature in London

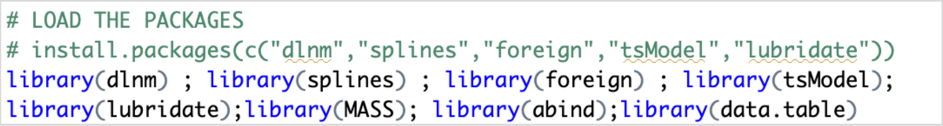
Lina Madaniyazi
10 July 2022

Introduction

This hands-on exercise describes the assessment of attributable risks due to air pollution and temperature in London. The document presents the basic steps of the analysis, such as assessing exposure-outcome associations, estimating attributable risks from forward and backward perspectives, separating attributable components, and computing uncertainty intervals. This illustration is based on a real-data case study of an analysis of the health burden due to air pollution and temperature in London using a time series dataset with daily observations from 2002 to 2006.

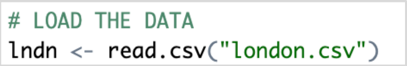
Date preparation

We start this demo by loading a few packages that provide useful functions for the examples:



```
# LOAD THE PACKAGES
# install.packages(c("dlnm","splines","foreign","tsModel","lubridate"))
library(dlnm) ; library(splines) ; library(foreign) ; library(tsModel);
library(lubridate);library(MASS); library(abind);library(data.table)
```

In this illustrate, we assess the mortality attributable to PM_{10} from a time series regression with a linear function for PM_{10}, and then estimate the mortality attributable to non-optimum temperature from a distributed lag non-linear model in London. We will make use of a real dataset, available in the CSV file **london**.

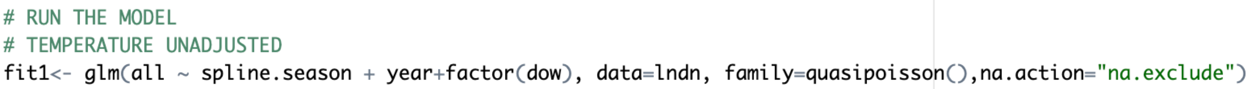


```
# LOAD THE DATA
lndn <- read.csv("london.csv")
```

The dataset includes the following variables:

- Date: date specified in Character format
- year: year-of-date
- month: month-of-date
- day: day-of-month, ranging from 1 to 31
- dow: day-of-week, ranging from 1 to 7
- all: counts of daily deaths for all-cause
- tmean: daily mean temperature (in ℃)
- rh: daily relative humidity (in %)
- PM10 and ozone: daily levels of pollution indices, namely inhalable coarse particles (PM_{10}, in $\mu g/m^3$) and ozone (in $\mu g/m^3$)

We first create the date variable:



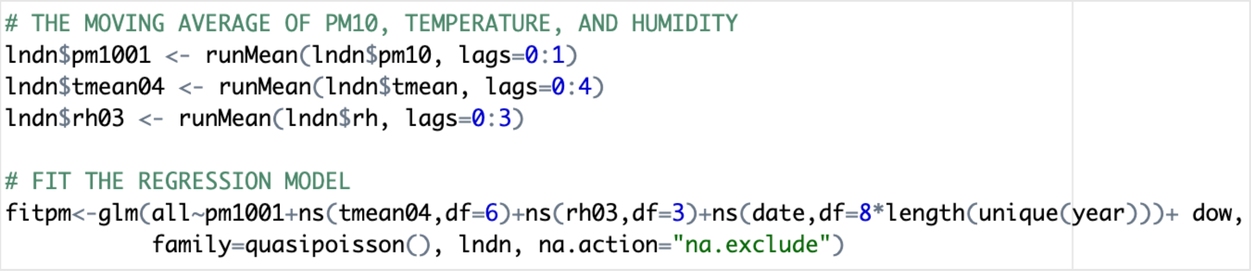
```
# RUN THE MODEL
# TEMPERATURE UNADJUSTED
fit1<- glm(all ~ spline.season + year+factor(dow), data=lndn, family=quasipoisson(),na.action="na.exclude")
```

Hands-on exercise 1: Attributable risk due to short-term exposure to PM$_{10}$

Assess the short-term association between PM$_{10}$ and mortality.

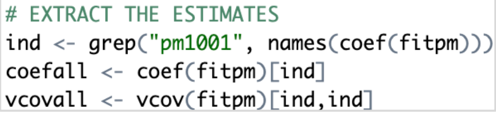
We estimate the association between PM10 and mortality using quasi-Poisson generalized additive model. We compute the moving average of daily PM10 concentration over lag 0-1 day and assumed a linear association. We include an indicator for the day-of-week to adjust for the within-week variation and a natural spline function with 8 degrees of freedom (df) per year to control for long-term trend and seasonality. We used two natural cubic spline functions with 6 df for the 4-day moving average of daily mean temperature and 3 df for the 3-day moving average of daily relative humidity.

Here is the code:



```r
# THE MOVING AVERAGE OF PM10, TEMPERATURE, AND HUMIDITY
lndn$pm1001 <- runMean(lndn$pm10, lags=0:1)
lndn$tmean04 <- runMean(lndn$tmean, lags=0:4)
lndn$rh03 <- runMean(lndn$rh, lags=0:3)

# FIT THE REGRESSION MODEL
fitpm<-glm(all~pm1001+ns(tmean04,df=6)+ns(rh03,df=3)+ns(date,df=8*length(unique(year)))+ dow,
           family=quasipoisson(), lndn, na.action="na.exclude")
```

To drive the attributable risk, we need to extract the coefficient (β) representing the short-term association between PM$_{10}$ and mortality ($AF_x = 1 - \exp(-\beta_x)$; $AN_x = n \cdot AF_x$).

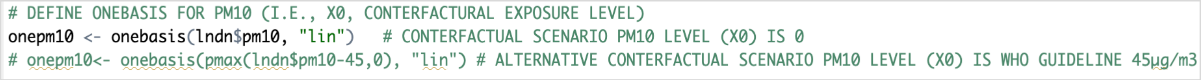


```r
# EXTRACT THE ESTIMATES
ind <- grep("pm1001", names(coef(fitpm)))
coefall <- coef(fitpm)[ind]
vcovall <- vcov(fitpm)[ind,ind]
```

Q1: What is your interpretation of the estimates (coefall) above?

Set the counterfactual condition.

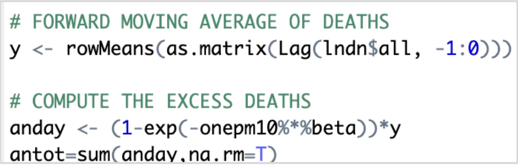
The theoretical nature of attributable risks is based on the definition and comparison of factual and counterfactual conditions. In other words, the observed condition is compared with a reference level, assuming that the same population is followed in an identical situation where only the exposure level changes to the reference value. Here, we define the counterfactual condition at $0\mu g/m^3$:



```r
# DEFINE ONEBASIS FOR PM10 (I.E., X0, CONTERFACTURAL EXPOSURE LEVEL)
onepm10 <- onebasis(lndn$pm10, "lin")   # CONTERFACTUAL SCENARIO PM10 LEVEL (X0) IS 0
# onepm10<- onebasis(pmax(lndn$pm10-45,0), "lin") # ALTERNATIVE CONTERFACTUAL SCENARIO PM10 LEVEL (X0) IS WHO GUIDELINE 45µg/m3
```

Assess attributable risks due to PM$_{10}$

In time series, we can compute the number of outcomes at time t attributable to an environmental stressor using the corresponding risk estimate associated with the exposure level at time t, by accounting for complex temporal patterns. We can drive the estimates from forward or backward perspectives. In the illustration, we estimate the association of mortality with 2-day moving average of PM10, representing the overall/cumulative effect of PM10 over lag 0-1. Since the lag-response association is impossible to derive from the regression model above directly, we derive the attributable risk from forward perspective. The *n* in the calculation of AN is derived by averaging the total counts experienced in the next t to t+lag (in this case, lag =1), approximating the lag structure of risks.

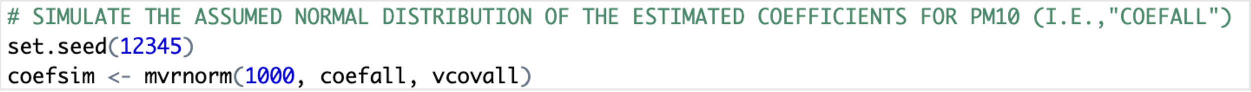


```
# FORWARD MOVING AVERAGE OF DEATHS
y <- rowMeans(as.matrix(Lag(lndn$all, -1:0)))

# COMPUTE THE EXCESS DEATHS
anday <- (1-exp(-onepm10%*%beta))*y
antot=sum(anday,na.rm=T)
```

Construct uncertainty interval

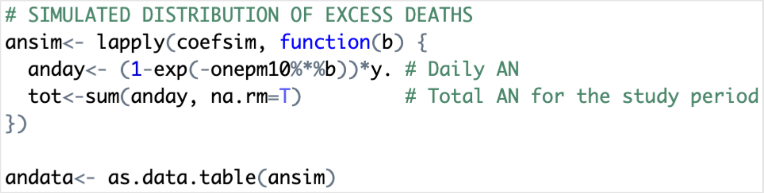
It is not easy to obtain the confidence intervals for attributable risks. Although several approximated estimators have been suggested, the most straightforward approach is to obtain interval estimation empirically through Monte Carlo simulations. Next, we use Monte Carlo simulations to compute 95% empirical confidence intervals (95%eCI) for AN and AF.

In specific, we simulate the assumed normal distribution of the estimated coefficients for PM10-mortality association and take random samples from the distribution.



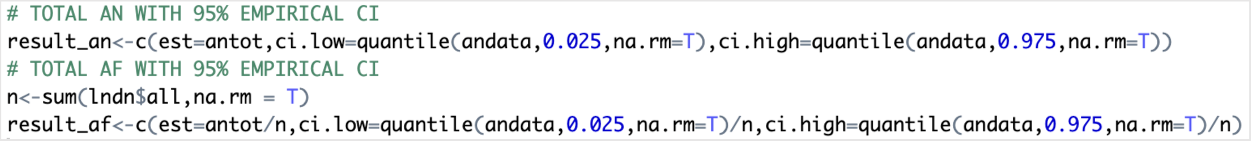
```
# SIMULATE THE ASSUMED NORMAL DISTRIBUTION OF THE ESTIMATED COEFFICIENTS FOR PM10 (I.E.,"COEFALL")
set.seed(12345)
coefsim <- mvrnorm(1000, coefall, vcovall)
```

These samples (i.e., “coefsim” above) are used to compute the attributable risks (“ansim/andata” below), empirically reconstructing the distributions of the attributable measures.



```
# SIMULATED DISTRIBUTION OF EXCESS DEATHS
ansim<- lapply(coefsim, function(b) {
  anday<- (1-exp(-onepm10%*%b))*y. # Daily AN
  tot<-sum(anday, na.rm=T)         # Total AN for the study period
})

andata<- as.data.table(ansim)
```

The related 2.5th and 97.5th percentiles of the distributions are interpreted as the 95% empirical confidence intervals (eCI).



```
# TOTAL AN WITH 95% EMPIRICAL CI
result_an<-c(est=antot,ci.low=quantile(andata,0.025,na.rm=T),ci.high=quantile(andata,0.975,na.rm=T))
# TOTAL AF WITH 95% EMPIRICAL CI
n<-sum(lndn$all,na.rm = T)
result_af<-c(est=antot/n,ci.low=quantile(andata,0.025,na.rm=T)/n,ci.high=quantile(andata,0.975,na.rm=T)/n)
```

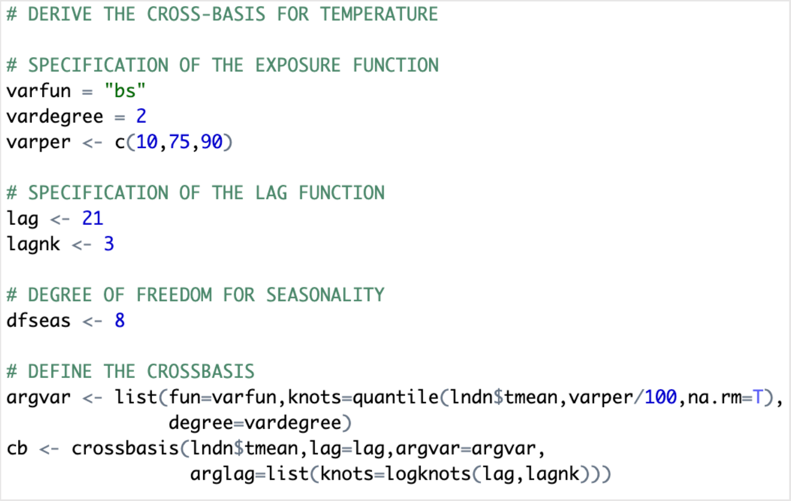
Q2: What is the attributable risk of PM10 to mortality in London between 2002 and 2006?

Hands-on exercise 2: Attributable risk due to short-term exposure to non-optimum temperature

A distributed lag non-linear model (DLNM) is used to assess the non-linear and lagged association between temperature and mortality, where the risk flexibly varies along a bi-dimensional space of the predictor and lag. Here, we use an extended definitions of attributable risk accounting for the additional temporal dimension from DLNM to assess the impact of non-optimum temperature on mortality in London.

Derive the cross-basis function for temperature

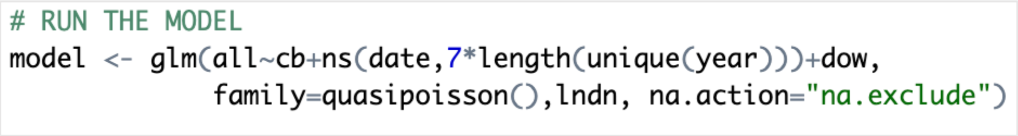
The cross-basis is composed of a quadratic B-spline with three internal knots placed at the 10th, 75th, and 90th percentiles of temperature distribution, and a natural cubic B-spline with an intercept and three internal knots placed at equally spaced values in the log scale over lags 0-21.



```
# DERIVE THE CROSS-BASIS FOR TEMPERATURE

# SPECIFICATION OF THE EXPOSURE FUNCTION
varfun = "bs"
vardegree = 2
varper <- c(10,75,90)

# SPECIFICATION OF THE LAG FUNCTION
lag <- 21
lagnk <- 3

# DEGREE OF FREEDOM FOR SEASONALITY
dfseas <- 8

# DEFINE THE CROSSBASIS
argvar <- list(fun=varfun,knots=quantile(lndn$tmean,varper/100,na.rm=T),
               degree=vardegree)
cb <- crossbasis(lndn$tmean,lag=lag,argvar=argvar,
                 arglag=list(knots=logknots(lag,lagnk)))
```

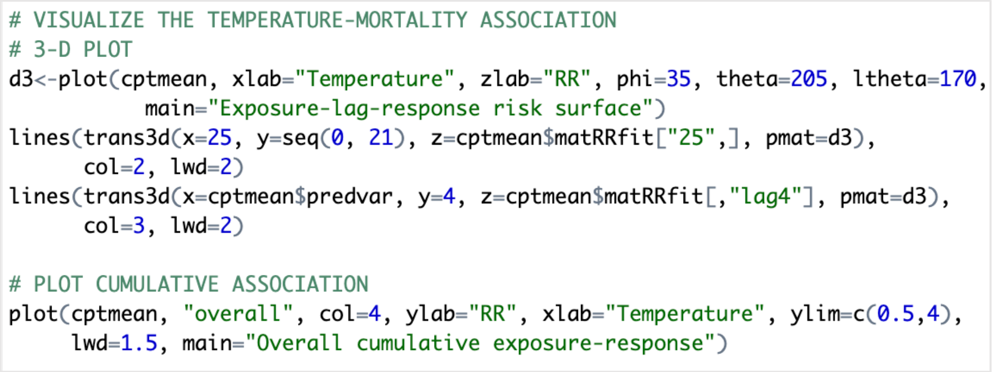
Run the model

We apply a time-series quasi-Poisson regression to estimate the delayed and non-linear association between temperature and mortality. The model includes A natural cubic B-spline of time with 8 df per year to control for seasonal and long-term trends and an indicator the day-of-week.



```
# RUN THE MODEL
model <- glm(all~cb+ns(date,7*length(unique(year)))+dow,
             family=quasipoisson(),lndn, na.action="na.exclude")
```

Visualize temperature-mortality association.

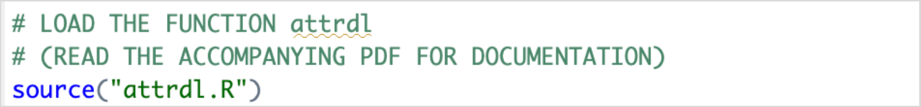


```
# VISUALIZE THE TEMPERATURE-MORTALITY ASSOCIATION
# 3-D PLOT
d3<-plot(cptmean, xlab="Temperature", zlab="RR", phi=35, theta=205, ltheta=170,
           main="Exposure-lag-response risk surface")
lines(trans3d(x=25, y=seq(0, 21), z=cptmean$matRRfit["25",], pmat=d3),
      col=2, lwd=2)
lines(trans3d(x=cptmean$predvar, y=4, z=cptmean$matRRfit[,"lag4"], pmat=d3),
      col=3, lwd=2)

# PLOT CUMULATIVE ASSOCIATION
plot(cptmean, "overall", col=4, ylab="RR", xlab="Temperature", ylim=c(0.5,4),
     lwd=1.5, main="Overall cumulative exposure-response")
```

Q3: What can you learn from the figures?

Assess attributable risks from DLNM

We load the attrdl.R function developed by Dr. Antonio Gasparrini to assess the attributable risks from backward and forward perspectives.

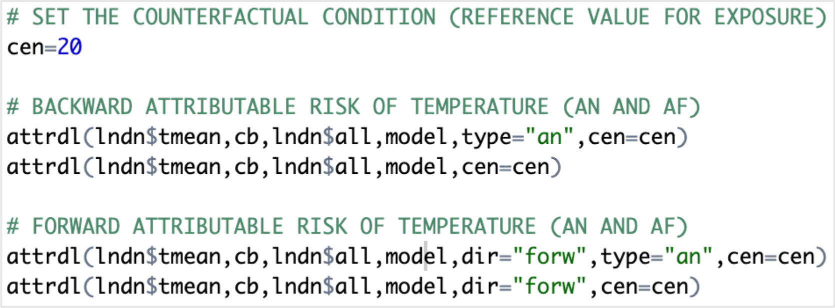


```
# LOAD THE FUNCTION attrdl
# (READ THE ACCOMPANYING PDF FOR DOCUMENTATION)
source("attrdl.R")
```

```
################################################################################
# SEE THE PDF WITH A DETAILED DOCUMENTATION AT www.ag-myresearch.com
#
#   - x: AN EXPOSURE VECTOR OR (ONLY FOR dir="back") A MATRIX OF LAGGED EXPOSURES
#   - basis: THE CROSS-BASIS COMPUTED FROM x
#   - cases: THE CASES VECTOR OR (ONLY FOR dir="forw") THE MATRIX OF FUTURE CASES
#   - model: THE FITTED MODEL
#   - coef, vcov: COEF AND VCOV FOR basis IF model IS NOT PROVIDED
#   - model.link: LINK FUNCTION IF model IS NOT PROVIDED
#   - type: EITHER "an" OR "af" FOR ATTRIBUTABLE NUMBER OR FRACTION
#   - dir: EITHER "back" OR "forw" FOR BACKWARD OR FORWARD PERSPECTIVES
#   - tot: IF TRUE, THE TOTAL ATTRIBUTABLE RISK IS COMPUTED
#   - cen: THE REFERENCE VALUE USED AS COUNTERFACTUAL SCENARIO
#   - range: THE RANGE OF EXPOSURE. IF NULL, THE WHOLE RANGE IS USED
#   - sim: IF SIMULATION SAMPLES SHOULD BE RETURNED. ONLY FOR tot=TRUE
#   - nsim: NUMBER OF SIMULATION SAMPLES
################################################################################
attrdl <- function(x,basis,cases,model=NULL,coef=NULL,vcov=NULL,model.link=NULL,
  type="af",dir="back",tot=TRUE,cen,range=NULL,sim=FALSE,nsim=1000)
################################################################################
```

Set the counterfactual condition

The theoretical nature of attributable risk is based on a counterfactual, where the observed condition is compared with a reference level, assuming that the same population is followed in an identical situation where only the exposure level changes to the reference value. Here, we set the reference level at 20°C and calculate AN and AF from backward and forward perspectives.

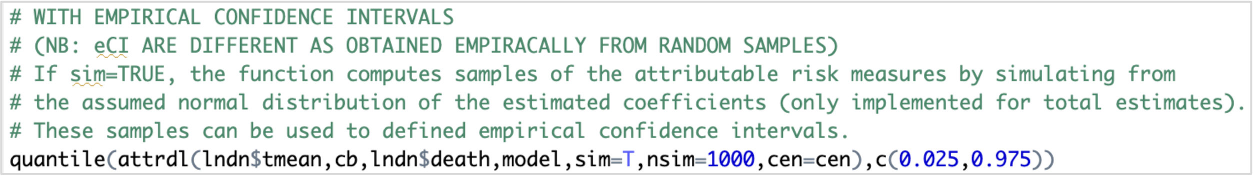


```
# SET THE COUNTERFACTUAL CONDITION (REFERENCE VALUE FOR EXPOSURE)
cen=20

# BACKWARD ATTRIBUTABLE RISK OF TEMPERATURE (AN AND AF)
attrdl(lndn$tmean,cb,lndn$all,model,type="an",cen=cen)
attrdl(lndn$tmean,cb,lndn$all,model,cen=cen)

# FORWARD ATTRIBUTABLE RISK OF TEMPERATURE (AN AND AF)
attrdl(lndn$tmean,cb,lndn$all,model,dir="forw",type="an",cen=cen)
attrdl(lndn$tmean,cb,lndn$all,model,dir="forw",cen=cen)
```

Q4: What can you learn from the results?

Construct uncertainty interval

It is not easy to obtain the confidence intervals for attributable risks. Although several approximated estimators have been suggested, the most straightforward approach is to obtain interval estimation empirically through Monte Carlo simulations.

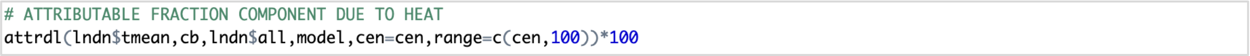


```
# WITH EMPIRICAL CONFIDENCE INTERVALS
# (NB: eCI ARE DIFFERENT AS OBTAINED EMPIRACALLY FROM RANDOM SAMPLES)
# If sim=TRUE, the function computes samples of the attributable risk measures by simulating from
# the assumed normal distribution of the estimated coefficients (only implemented for total estimates).
# These samples can be used to defined empirical confidence intervals.
quantile(attrdl(lndn$tmean,cb,lndn$death,model,sim=T,nsim=1000,cen=cen),c(0.025,0.975))
```

Separate attributable components

We can further separate the attributable components related to heat and cold. Although this computation can be done from forward and backward perspectives, the forward version is well suited for separating attributable components, as their sum matches the overall risk.

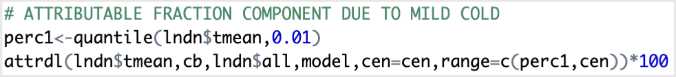
Here, we derive the AF component due to heat, which is defined as temperature ranging from the reference level (20°C) to 100°C. You may change the upper limit to any level which is above the maximum of daily mean temperature in the data.



```
# ATTRIBUTABLE FRACTION COMPONENT DUE TO HEAT
attrdl(lndn$tmean,cb,lndn$all,model,cen=cen,range=c(cen,100))*100
```

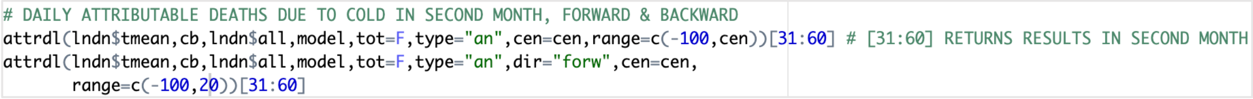
Q5: Can you derive AN and AF due to heat with a reference level at 25℃? Is it the same with the results above?

Next, we derive AF component due to mild cold, which is defined as the temperature ranging from daily mean temperature at the 1st percentile to the reference level (20°C)



```
# ATTRIBUTABLE FRACTION COMPONENT DUE TO MILD COLD
perc1<-quantile(lndn$tmean,0.01)
attrdl(lndn$tmean,cb,lndn$all,model,cen=cen,range=c(perc1,cen))*100
```

At last, let’s calculate AN due to cold in the second month from backward and forward perspectives. Here, we need to set “tot=F”, in order to obtain daily AN.



```
# DAILY ATTRIBUTABLE DEATHS DUE TO COLD IN SECOND MONTH, FORWARD & BACKWARD
attrdl(lndn$tmean,cb,lndn$all,model,tot=F,type="an",cen=cen,range=c(-100,cen))[31:60] # [31:60] RETURNS RESULTS IN SECOND MONTH
attrdl(lndn$tmean,cb,lndn$all,model,tot=F,type="an",dir="forw",cen=cen,
       range=c(-100,20))[31:60]
```

References:

- Gasparrini A, Leone M. Attributable risk from distributed lag models. BMC Medical Research Methodology. 2014;14(1):55.
- Kyle Steenladn and Ben Armstrong. An overview of methods for calculating the burden of disease due to specific risk factors. Epidemiology. 2006;17:512-519.
- Vicedo-Cabrera AM, Sera F, Gasparrini A. Hands-on tutorial on a modeling framework for projections of climate change impacts on health. Epidemiology. 2019;30(3):321-329.
- O'Brien E, Masselot P, Sera F, Roye D, Breitner S, Ng CFS, de Sousa Zanotti Stagliorio Coelho M, Madureira J, Tobias A, Vicedo-Cabrera AM, Bell ML, Lavigne E, Kan H, Gasparrini A; MCC Collaborative Research Network. Short-term association between sulfur dioxide and mortality: a multicountry analysis in 399 cities. Environmental Health Perspectives. 2023;131(3):37002.

R code and other references:

- www.ag-myresearch.com/r-code
- github.com/gasparrini