Causal effects of air pollution on mortality: investigating the role of temperature

Laurea magistrale in Statistica e Data Science
Academic year 2022/2023

Thesis advisor: Michela Baccini
Candidate: Andrea Marchetti

UNIVERSITÀ
DEGLI STUDI
FIRENZE

Climate change is one of the most pressing challenges of the 21*st* century and is closely linked to rising **air pollution** and **temperatures**.

Since the 90*s*, many studies have found increased mortality and hospitalisation due to exposure to air pollutants and extreme/moderately extreme temperatures.

Although these studies have examined the effects of pollution or temperature on **mortality**, evidence on their combined is still limited in epidemiological research.

The aim of this work is to investigate, within a causal inference framework, how temperature modifies the health effects of fine particles and how these two exposures interact with each other. We pursue this goal by resorting to *g-computation*, on data from the city of Milan.

Robins' g-computation allows us to answer the question *"what would happen if...?"*, describing how the joint probability distribution of the response variable would change if we hypothetically intervened on one or more exposure/treatment variables.

If we are interested in comparing two different interventions $(Z = 0, Z = 1)$, we focus on the **causal marginal mean difference** given by the expected value of the difference between the two potential outcomes $Y(1)$ and $Y(0)$ (SUTVA, consistency):

$$E[Y(1)-Y(0)] = \sum_{\boldsymbol{x}} \{E(Y|Z=1, \boldsymbol{X}=\boldsymbol{x}) - E(Y|Z=0, \boldsymbol{X}=\boldsymbol{x})\} Pr(\boldsymbol{X}=\boldsymbol{x})$$

where $\boldsymbol{X}$ is a vector of relevant covariates.

This difference can be obtained by averaging over the predicted potential outcomes obtained for each observation at each level of treatment(s) from a *Q-model*, i.e. a regression model for the observed outcome given the observed exposures and covariates.

(VanderWeele 2009, Keele and Stevenson 2017)

Definition: EFFECT MODIFICATION

A variable T is said to be an effect modifier on the causal risk difference scale for the effect of P on D if T is not an effect of P and if there are 2 levels of P, p_0 ed p_1, and 2 levels of T, t_0 e t_1, such that

$$E[D(p_1)|T=t_1]-E[D(p_0)|T=t_1] \neq E[D(p_1)|T=t_0]-E[D(p_0)|T=t_0]$$

Definition: INTERACTION EFFECT

There is said to be an interaction on the causal risk difference scale between the effects of P and T on D, if there are 2 levels of P, p_0 ed p_1, and 2 levels of T, t_0 e t_1, such that

$$E[D(p_1,t_1)] - E[D(p_0,t_1)] \neq E[D(p_1,t_0)] - E[D(p_0,t_0)].$$

In this work, we evaluate both effects on the scale of percent changes, in the case of continuous treatment(s).

We examined health data, air pollution data, and daily environmental data from Milan, collected from the air quality monitoring stations of the ARPA Lombardia and the Regional Mortality Register. This analysis covers four years, from 2003 to 2006, encompassing a total of 1461 days.

Let $i = 1, ..., N$ be the indicator of the day. We denote by:

- $P_i \in \mathcal{P}$ the PM_{10} exposure level on day i (lag exposure 0-1);
- $T_i \in \mathcal{T}$ the temperature on day i (lag exposure 0-3);
- $D_i \in \mathcal{D}$ the number of deaths on day i;
- $\boldsymbol{X}_i \in \mathcal{X}$ a vector of K covariates for day i.

To assess the effect of the treatment(s), we make the following assumptions:

1. SUTVA that allows us to define
   - $\boldsymbol{D_i(p)}$ the potential number of deaths on day i if p were the exposure level on that day;
   - $\boldsymbol{D_i(p,t)}$ the potential number of deaths on day i if the PM_{10} exposure level were p and a temperature were t.
2. Unconfoundedness

$$D(p) \perp\!\!\!\perp P|\boldsymbol{X} \quad or \quad D(p,t) \perp\!\!\!\perp P,T|\boldsymbol{X}.$$

3. Positivity

$$0 < Pr(P_i = p|\boldsymbol{X}_i) < 1 \quad or \quad 0 < Pr(P_i = p, T_i = t|\boldsymbol{X}_i) < 1 \quad \forall i.$$

Identification of the causal interaction requires validity of the previous assumptions for both treatments, whereas the effect modification requires it only for the primary treatment.

To predict potential outcomes for each observation at each level of exposure to the treatment or treatments, we specified the following *Q-model*:

$$D_i \mid P_i, T_i, \boldsymbol{X}_i \sim quasi - Poisson(\lambda_i)$$

where $log(\lambda_i)$ is explained by:

- indicators for day of week, summer season and flu epidemic;
- penalised linear regressions spline for seasonality, humidity, temperature and PM_{10};
- two tensor products between temperature and humidity and between temperature and PM_{10}.

NOTE: that for the interaction we specified separated models by season, in order to guarantee positivity.

We estimated the following quantities:

- *average dose-response function* (aDRF)

$$\widehat{\mu}(p) = \frac{1}{N}\sum_{i=1}^{N} \widehat{D}_i(p)$$

- *conditional average dose-response functions*

$$\widehat{\mu}(p, \boldsymbol{x}) = \frac{1}{N(\mathcal{X}^*)}\sum_{i:\boldsymbol{X}\in\mathcal{X}^*} \widehat{D}_i(p)$$

- *average dose response surface*

$$\widehat{\mu}(p, t) = \frac{1}{N}\sum_{i=1}^{N} \widehat{D}_i(p, t)$$

A non-parametric bootstrap over 1000 samples was implemented to obtain the 90% pointwise confidence bands.

Causal effect of PM_{10} on mortality

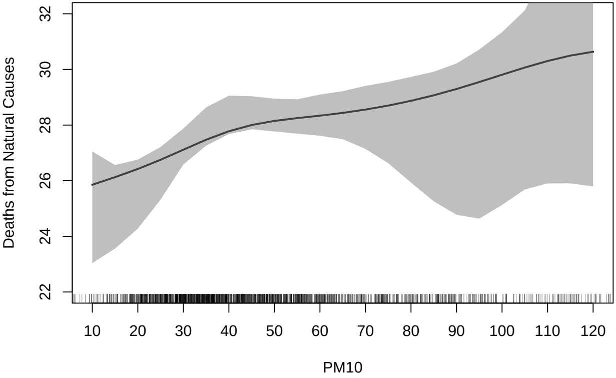

Figure: Average Dose-Response Function describing the causal relationship between PM_{10} exposure at lag 0-1 and average daily mortality from natural causes (90% pointwise confidence band).

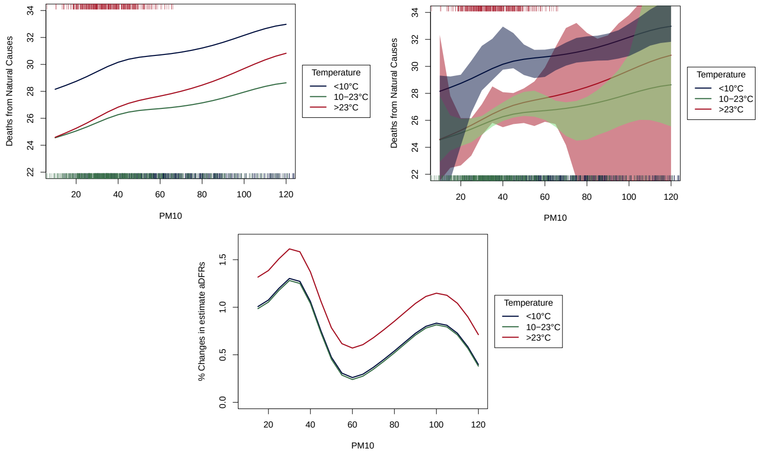

Figure: aDRF describing the causal relationship between PM_{10} exposure and daily mortality from natural causes, and related percent changes associated to an increases for 5 $\mu g/m^3$ of PM_{10}, by temperature categories (90% point-wise confidence bands).

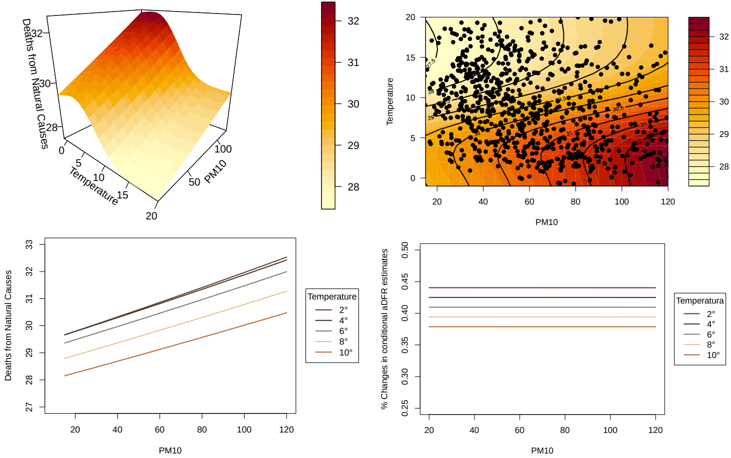

Figure: Average DR surface and contour plot describing the joint effect of PM_{10} and temperature on mortality, conditional average DRF describing the effect of PM_{10} on mortality by temperature and associated percent changes for 5 $\mu g/m^3$ of PM_{10} - cold season.

UNIVERSITÀ DEGLI STUDI FIRENZE

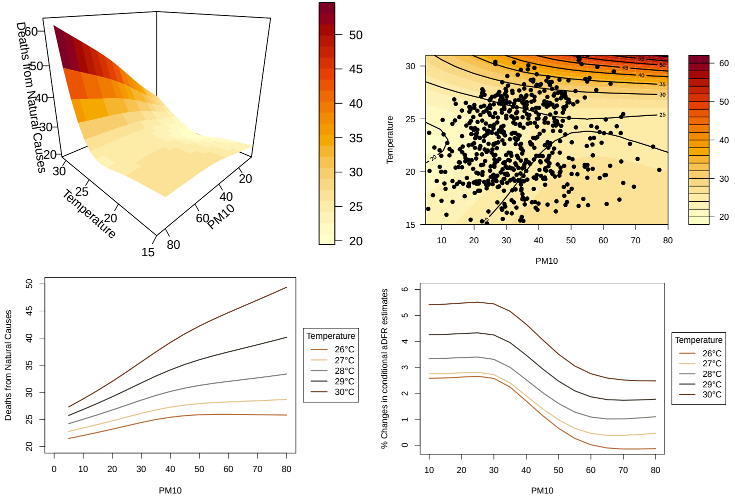

Figure: Average DR surface and contour plot describing the joint effect of PM_{10} and temperature on mortality, conditional average DRF describing the effect of PM_{10} on mortality by temperature and associated percent changes for 5 $\mu g/m^3$ of PM_{10} - warm season.

- We used *g-computation* to derive an estimate of the average dose-response function/surface that describes the short term causal effect of PM_{10} on mortality by temperature, both considering temperature as an effect modifier and considering the interaction of the two exposures.
- Our study confirms the causal relationship between air pollution and mortality, and suggests the *clear existence of effect modification by temperature and interaction with it.*
- The two approaches led to qualitatively similar results. We must account that the research question is different in the two cases as well as the reported estimands.
- Our research suggests that public health responses to climate change should take into account the *combined health impacts* of air pollution and temperature.
- *Limitations:* the approach requires the correct specification of the Q-model; possible violation of the positivity assumption.

Thank you
for your attention