CAUSAL INFERENCE ON THE IMPACT OF EXTREME AMBIENT TEMPERATURES ON POPULATION HEALTH

Michela Baccini[1], Alessandra Mattei[1], Elena Degli Innocenti[1], Giulio Biscardi[1], Aitana Lertxundi[2,3]

[1] Department of Statistics, Computer Science, Applications, University of Florence, Italy
(e-mail: michela.baccini@unifi.it)

[2] Department of Preventive Medicine and Public Health, University of Basque Country, Leioa, Spain

[3] Biodonostia Health Research Institute, San Sebastian, Spain

ABSTRACT: A potential outcome approach to causal inference is used to infer the average exposure-response curve describing the relationship between daily temperature and daily mortality in the city of San Sebastian (Spain) for the period 2010-2015. The analysis relies on the estimate of the generalized propensity score and specification of a model for potential outcomes. The impact of extreme temperatures on population health is also provided, in terms of attributable deaths.

KEYWORDS: temperature, mortality, dose-response curve, generalized propensity score, health impact assessment.

1 Introduction

Climate change is now regarded as the greatest challenge of the 21st century. Extreme temperature levels are one of its consequences. Many studies, based on the analysis of daily time series through regression approaches, have identified a U-, V- or J-shaped relationship between environmental temperature and mortality, indicating that heat and cold are associated with death counts. For the first time, this study estimates this relationship by using a potential outcome approach to causal inference. The method proposed is based on the generalized propensity score and uses a semi-parametric specification for the outcome model. We ground on the method used in Forastiere et al. (2020) for the analysis of the short term effect of air pollution on mortality in the city of Milan (Italy).

2 Data

The health and exposure data used in this paper have been collected for the city of San Sebastian (Basque Country region of Spain) for the period 2010-2015. They include the daily number of deaths from natural causes, cardiovascular and respiratory causes, grouped by age (0-64, 65-84, 85+); meteorological variables (temperature and humidity); and several known confounders of the temperature-mortality relationship (average pollutant levels and an indicator of influenza epidemics).

3 Methods

The analyses are performed separately for cold and warm season.
According to the potential outcome framework, under the Stable Unit Treatment Value Assumption (Imbens and Rubin 2015), we denote by $Y_i(z)$ the potential number of deaths in day i ($i = 1, 2, \ldots, n$) if z were the level of temperature in that day. For each day we only observe one potential outcome, that is, the one corresponding to the actual exposure of that day, Z_i, all the other potential outcomes with $z \neq Z_i$ being missing. We denote the observed outcome with Y_i, while $\mathbf{X}_i = (x_{1i}, x_{2i}, \ldots, x_{Ki})$ is the vector of the K covariates measured on day i.
We are interested in the average Dose Response Function (aDRF), defined as:

$$\mu(z)=n^{-1}\sum_i Y_i(z). \qquad [1]$$

Under the unconfoundedness assumption, we fill in missing potential outcomes in [1] following the procedure described in Hirano and Imbens (2004), which requires the specification of a model for the exposure, used for GPS estimation, and a model for the outcome.
The model for the exposure is a log-Normal model on the daily average temperature Z_i, given the confounders ($\mathbf{X}_i$) and seasonality terms. The confounders are included in the model through flexible functions and interactions are allowed. The GPS for day i at the level of exposure z is then defined as the value of the density function for $\log(Z)$, derived from the estimated model:

$$r(z, \mathbf{X}_i)=(2\pi s)^{-1}\exp[-(\log(z)-m_i)^2/(2s^2)],$$

where m_i is the value of $\log(Z)$ predicted by the model for day i, and s^2 is the estimated error variance.
The model for the outcome is a Poisson regression model on the daily mortality Y_i, given both daily average temperature Z_i and the value of GPS estimated for $z = Z_i$, $R_i = r(Z_i, \mathbf{X}_i)$. Different specifications of the outcome model can be adopted: we define a bivariate spline on temperature and GPS.
Once the two models have been estimated, there is the phase of prediction and potential outcome imputation. After defining a grid of temperatures, we calculate, for each day, the GPS on each value z^* of the grid. Then, we plug in z^* and the corresponding GPS, $r(z^*, \mathbf{X}_i)$, in the estimated outcome model, in order to predict the mortality level $Y_i(z^*)$ that would be observed if the temperature in day i were equal to z^*. Finally, for each z*, the predicted potential outcomes are averaged over the days, so that an exposure-response curve is obtained.

We estimate the aDRF for mortality from all causes and by cause of death, for all age and separately by age group. Also, we estimate, in terms of attributable deaths, the impact of temperatures higher or lower than specific thresholds on population mortality. Confidence intervals are obtained through a block-bootstrap procedure.
A crucial point in the analysis is the specification of the exposure model. The validity of the specification adopted for the exposure model is assessed by checking the covariate balance as described in Hirano and Imbens (2004).

4 Results

Extreme temperatures, both cold and warm, have a detrimental effect on health. The so-called `turning point', defined as the temperature where the aDRF is minimum, is found to be around 19.5° C. The analysis by age group confirms these effects for people over 65 years of age, while negligible effects are observed for younger people (0-64).
Taking the value of 19.5°C as an optimal threshold for health, we estimate that, in the warm season, exceeding it has caused 115 deaths (90% CI: 22.39, 229.31) during the study period. In the cold season, staying below the same threshold is estimated to have caused 483 deaths (90% CI: 97.21, 836.64).

5 Discussion

This study states the existence of a causal relationship between temperature and mortality and provides an approach to estimate the average dose-response function, as well as the impact of extreme exposures. Extensions of the method could allow the estimation of an entire curve on the whole year and the investigation of the delayed effect of temperature on mortality.

References

FORASTIERE, L., CARUGNO, M., & BACCINI, M. 2020. Assessing short-term impact of PM10 on mortality using a semiparametric generalized propensity score approach. *Env Health.* **19**, 46.

IMBENS, G., & RUBIN, D. 2015. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge: Cambridge University Press.

IRANO, K. & IMBENS, G.W. 2004. The propensity score with continuous treatments. In: Gelman, A., Meng, X.L., editors. Applied Bayesian Modelling and CausalInference from Missing Data Perspectives: An Essential Journey with Donald Rubin's Statistical Family. p. 73–84.