BY KATRIEN ANTONIO, TORSTEN KLEINOW AND JENS ROBBEN

The short-term association between environmental variables and mortality

Dr. J. Robben is postdoctoral researcher at the Research Centre for Longevity Risk of the University of Amsterdam. His current research explores diverse aspects of mortality and longevity modelling, with a focus on the impact of environmental and epidemiological variables on mortality, old-age mortality patterns, and the use of machine learning techniques in mortality modelling.

Prof. dr. K. Antonio is full professor in actuarial science at KU Leuven (Belgium) and part-time professor in actuarial data science with the University of Amsterdam (the Netherlands). Currently, she serves as vice-dean for education and students at the Faculty of Economics and Business (FEB), KU Leuven.

Prof. dr. T. Kleinow is Professor for Longevity Risk at the University of Amsterdam. His primary research interests include mortality modelling and the analysis of socio-economic disparities in life expectancy. He has published in international peer-reviewed journals and contributed to several international research projects on mortality and longevity.

Traditionally, stochastic mortality models used for actuarial purposes are built at the national level and on an annual time scale. While useful for long-term projections, these models provide little insight into how mortality fluctuates within a year. It is well known that mortality has a seasonal structure, where rates often peak in winter and are lowest during summer. Deviations from this seasonal structure may occur for several reasons, for example, due to extreme environmental conditions, such as heat waves, cold spells, and high air pollution. From a public health perspective, it is important to understand what drives such excess mortality and how we can properly quantify its extent, especially in the context of climate change. From an actuarial perspective, understanding these short-term mortality dynamics is essential in a life insurance context for the estimation of incurred-but-not-reported death benefit payments, pricing short-term life-contingent insurance products, and accurately assessing short-term fluctuations in mortality risk. Using machine learning techniques, we examine the drivers of short-term excess mortality among older adults and shed light on the impact of extreme weather conditions and air pollution.

DATA

We make use of several publicly available datasets. First, we extract Eurostat's weekly death counts and population exposures for the 65+ female age group in more than 550 European regions for the period 2013-2019. Second, we download daily weather data, i.e., maximum temperature, minimum temperature, relative humidity, total precipitation, and average wind speed from the Copernicus Climate Data Store. Third, we retrieve hourly average concentration levels for ozone, nitrogen dioxide, and particulate matter (PM10 and PM2.5) from the Copernicus Atmosphere Monitoring Service. Since our weekly mortality data does not align in temporal and geographical scale with the daily (or hourly) environmental data that is defined on a fine spatial grid, we perform an extensive feature engineering process to harmonise these datasets. More specifically, we construct features that reflect the severity and frequency of the environmental conditions within a week. For instance, for temperature, we construct features that measure the frequency of hot days in a week (hot-week index) by looking at the number of days a preselected high threshold is exceeded, as well as features that measure the severity of hot days within a week by looking at the average daily temperature anomalies in that week.

WE IDENTIFY THE KEY ENVIRONMENTAL DRIVERS OF DEVIATIONS FROM THE MORTALITY BASELINE

MODELLING APPROACH

We develop a two-stage modelling framework to identify short-term associations between environmental factors and weekly mortality fluctuations across the entire set of European regions. In the first stage, a regional baseline model estimates the expected number of weekly deaths in each region using a classical Poisson regression with Fourier terms to capture seasonal patterns. In the second stage, deviations from this baseline (excess or deficit mortality) are modelled using a machine learning algorithm, i.e., extreme gradient boosting or XGBoost. The model incorporates engineered features describing the severity and frequency of extreme environmental conditions within a week, their lagged values, geographic coordinates, and seasonal indicators. This approach automatically captures the complex, non-linear relationships and interactions between the different weather and air pollution features and their combined impact on weekly mortality. Using machine learning interpretation tools such as feature importance and accumulated local effects, we identify the key environmental drivers of deviations from the mortality baseline and quantify their marginal effects.

RESULTS AND INSIGHTS

After calibrating the two-stage model on the outlined mortality and environmental data, we find that short-term deviations from baseline mortality are primarily driven by temperature. XGBoost identifies the lagged weekly average of the daily minimum temperature anomalies, the hot-week index, and the lagged cold-week index as the most important features to explain these deviations, while wind speed and precipitation play only a minor role. Air pollution features have a smaller but statistically significant marginal impact.

We also find clear evidence of a heat-related harvesting effect: a hot week causes immediate excess mortality, followed by a mortality deficit in the subsequent week with normal temperatures. No harvesting effect is observed for cold weeks. In addition, we find that mortality increases when both PM10 levels and temperatures are high.

Regionally, we observe that environmental effects on mortality are stronger in southern Europe than in northern regions. To illustrate this, we compute the relative risk of mortality compared to the baseline for each considered European region. As an example, we focus on a typical week at the start of July and assume average air pollution concentrations, humidity, rainfall, and wind speed for that week in each region. We consider two heat scenarios. In the moderate heat scenario, the weekly average of the daily minimum and maximum temperature anomalies is 6°C, and half of the days surpass the preselected hot temperature threshold. In the severe heat scenario, the anomalies rise to 12°C (the maximum observed in our sample), and all days exceed the threshold. Using the calibrated XGBoost model, we then predict the relative risk, expressed as an amplification of baseline mortality under each scenario.

Figure 1 shows these results across European regions, highlighting clear differences between north and south. Under the severe heat scenario, mortality can increase by up to 25% in southern regions such as Spain, Portugal, and Italy, while the increase remains below 10% in northern regions.

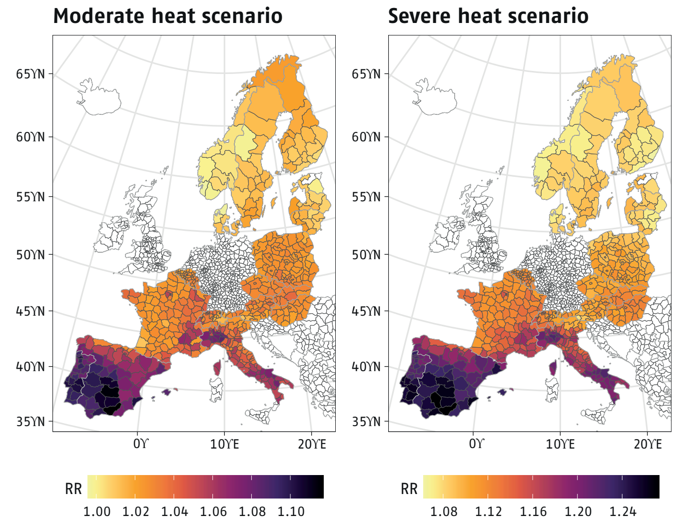

Figure 1: Relative risk by region and scenario

CONCLUSION

This study develops a two-stage modelling framework integrating environmental factors into weekly mortality models across 550+ European regions. Using a regional baseline and machine learning, we find that temperature extremes are the strongest drivers of short-term excess deaths, with clear regional disparities and evidence of harvesting effects. Some air pollution features also contribute, while other weather variables, such as rainfall and wind speed, are less important. The framework provides a useful tool to assess and quantify the environmental drivers of excess mortality, and hence, finds its use in both insurance and public health. ■

This article is based on the paper *Robben, J., Antonio, K., & Kleinow, T. (2025).* The short-term association between environmental variables and mortality: evidence from Europe. *Journal of the Royal Statistical Society: Series A (Statistics in Society), qnaf052.* (https://academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnaf052/8142302)