Data in Brief 55 (2024) 110694

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

# Compilation of open access time-series datasets for studying temperature-mortality association

Aurelio Tobias[a,b,*], Chris Fook Sheng Ng[c], Yoonhee Kim[d], Masahiro Hashizume[c], Lina Madaniyazi[b]

[a] *Institute of Environmental Assessment and Water Research (IDAEA), Spanish Research Council (CSIC), Barcelona, Spain*
[b] *School of Tropical Medicine and Global Health (TMGH), Nagasaki University, Nagasaki, Japan*
[c] *Department of Global Health Policy, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan*
[d] *Department of Global Environmental Health, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan*

A R T I C L E I N F O

*Article history:*
Received 30 April 2024
Revised 11 June 2024
Accepted 23 June 2024
Available online 28 June 2024

Dataset link: tsdatasets (Reference data)

*Keywords:*
Environmental health
Climate change
Environmental epidemiology
Time-series regression
Multi-location studies

A B S T R A C T

In this article, we present a comprehensive compilation of open access daily time-series datasets tailored to assess the temperature-mortality association. The data consists of daily mortality counts and average ambient temperature at various levels of geographic aggregation, including data from four cities, ten regions, and two counties, which have been utilised in previously published studies. These datasets are applicable for time-series regression analysis to estimate location-specific temperature-mortality associations. Additionally, the availability of data from multiple geographical locations enabled the exploration of geographical differences by pooling associations using meta-analysis. This compilation aims to serve as a valuable resource for researchers, educators, and students, facilitating their application of time-series regression modelling for research endeavours and training activities.

* Corresponding author.
*E-mail address:* aurelio.tobias@idaea.csic.es (A. Tobias).
*Social media:* @atobias69 (A. Tobias), @chrisngfs (C.F.S. Ng), @yoonhee326 (Y. Kim), @Lina_Japan (L. Madaniyazi)

https://doi.org/10.1016/j.dib.2024.110694

Specifications Table

| | |
|---|---|
| Subject | Environmental health. |
| Specific subject area | Temperature-mortality short-term association with time-series regression. |
| Type of data | Raw, Filtered |
| Data collection | All the variables were collected on a daily basis between 1987 and 2007. However, study periods may change between cities and countries. Daily counts of all-cause mortality, cardiovascular mortality, and respiratory mortality. All-cause mortality data was also collected by age group (<65, ≥65 years). The location-specific daily average ambient temperature (°C) was recorded. |
| Data source location | Baltimore (US), Chicago (US), Valencia (Spain), London (UK), 10 regions of England and Wales (UK), Greece. |
| Data accessibility | Repository name: Mendeley data<br>Data identification number: (10.17632/kjgrmbnt5h.4)<br>Direct URL to data: https://data.mendeley.com/datasets/kjgrmbnt5h/ |
| Related research article | The authors collected data from public Governmental data repositories accessible under request to be used in previously published studies [13]. |

## 1. Value of the Data

- This compilation of time-series datasets allows assessing the temperature-mortality association and its short-term health effects in multiple locations.
- The data can be used for educational purposes to illustrate the use of time-series regression and distributed lag non-linear models in temperature-mortality studies.
- The data provides a basis for understanding time-series regression modelling in environmental epidemiology studies, allowing for the reproducibility of results.

## 2. Background

Environmental epidemiology examines the relationship between environmental exposures and health outcomes in human populations. In this context, time-series regression studies have been widely used to investigate the short-term associations between environmental exposures (e.g., ambient temperature) and daily health outcomes (e.g., daily mortality counts) [1]. Time-series studies provide insights into lagged effects, thresholds, and potential effect modifications, which are valuable for policymaking and public health interventions. Advancements in statistical methods, particularly the application of distributed lag non-linear models (DLNMs) within time-series regression, have contributed to disentangling complex time-varying exposure-response relationships [2].

In recent years, the push for open science and data transparency has triggered the sharing of datasets across disciplines [3,4]. Open access datasets foster collaboration and reproducibility, and serve as teaching tools and exemplars for research. There is a compelling need to compile time-series datasets relevant to environmental epidemiology. We present a comprehensive compilation and description of open access time-series datasets tailored for studying the temperature-mortality association. This compilation aims to serve as a valuable resource for researchers, educators, and students, facilitating their application of time-series regression modelling for research endeavours and training activities.

## 3. Data Description

### *3.1. Setting*

We compiled daily time-series data at the city level for Baltimore (US) [5] and Chicago (US) between 1987 and 2000 [6], and for Valencia (Spain) between 2001 and 2008 [7]; at the regional level for the London metropolitan region and other regions of England and Wales (UK) between 1993 and 2012, [8-10]; and at the country level for England and Wales between 1993 and 2012, and for Greece between 2001 and 2007 [11,12].

### *3.2. Mortality data*

Mortality data are represented by daily counts for all causes, excluding external causes (natural mortality, International Classification of Diseases, 9th and 10th Revisions, ICD-9: 1–799 and ICD-10: A00-R99), cardiovascular diseases (ICD-9: 390–459, ICD-10: I00-I99), and respiratory diseases (ICD-9: 460–519, ICD-10: J00-J99). Mortality counts for all causes are available for all locations, while counts for cardiovascular and respiratory diseases are only available in Baltimore, Chicago, Valencia, and London. Additionally, mortality counts for all causes were classified into two age groups ($<65$, $\geq 65$ years), which are commonly used in environmental epidemiology studies to identify vulnerable age groups, in Baltimore, Valencia, and London. All series of daily mortality counts were complete (no missing values).

Data were obtained from the National Center for Health Statistics (Baltimore, Chicago), the Valencian Community Mortality Register (Valencia), the National Center for Health Statistics (London, regions of England and Wales), and the Hellenic Ministry of the Interior (Greece).

### *3.3. Temperature data*

Daily mean ambient temperature (°C) was collected for all locations from the National Climatic Data Center (Baltimore, Chicago), the Spanish National Meteorological Agency (Valencia), and the National Oceanic and Atmospheric Administration (London, regions of England and Wales, Greece). All series of temperature data were complete (no missing values).

## 4. Experimental Design, Materials and Methods

The compiled datasets have been used in previous studies to evaluate the temperature-mortality association using time-series regression [13]. This illustrative example applies a two-stage design to model multi-location data [14] using the time-series datasets for Baltimore, Chicago, Valencia, and London (Fig. 1). Statistical analysis was performed in R software, version 4.3.3. The R code for analising the example is available in the Appendix as Supplementary Data.

In the first stage, we analysed city-specific temperature-mortality associations using quasi-Poisson regression with DLNMs [2]. This class of models can describe complex non-linear and lagged dependencies by combining two functions: one for the exposure-response association and another for the lag-response association. For each location, we modelled the temperature-mortality relationship using a natural cubic spline with three internal knots at the 10th, 75th, and 90th percentiles of the temperature distribution, and the lag-response relationship using a natural cubic spline with three internal knots equally spaced on the logarithmic scale. The lag period was extended to 21 days to capture the long delay in the effects of cold. The model also included a natural cubic spline of time with 10 degrees of freedom per year to for control seasonal variations and long-term trends, and indicator variables for days of the week. These modelling choices are based on extensive previous work [13].

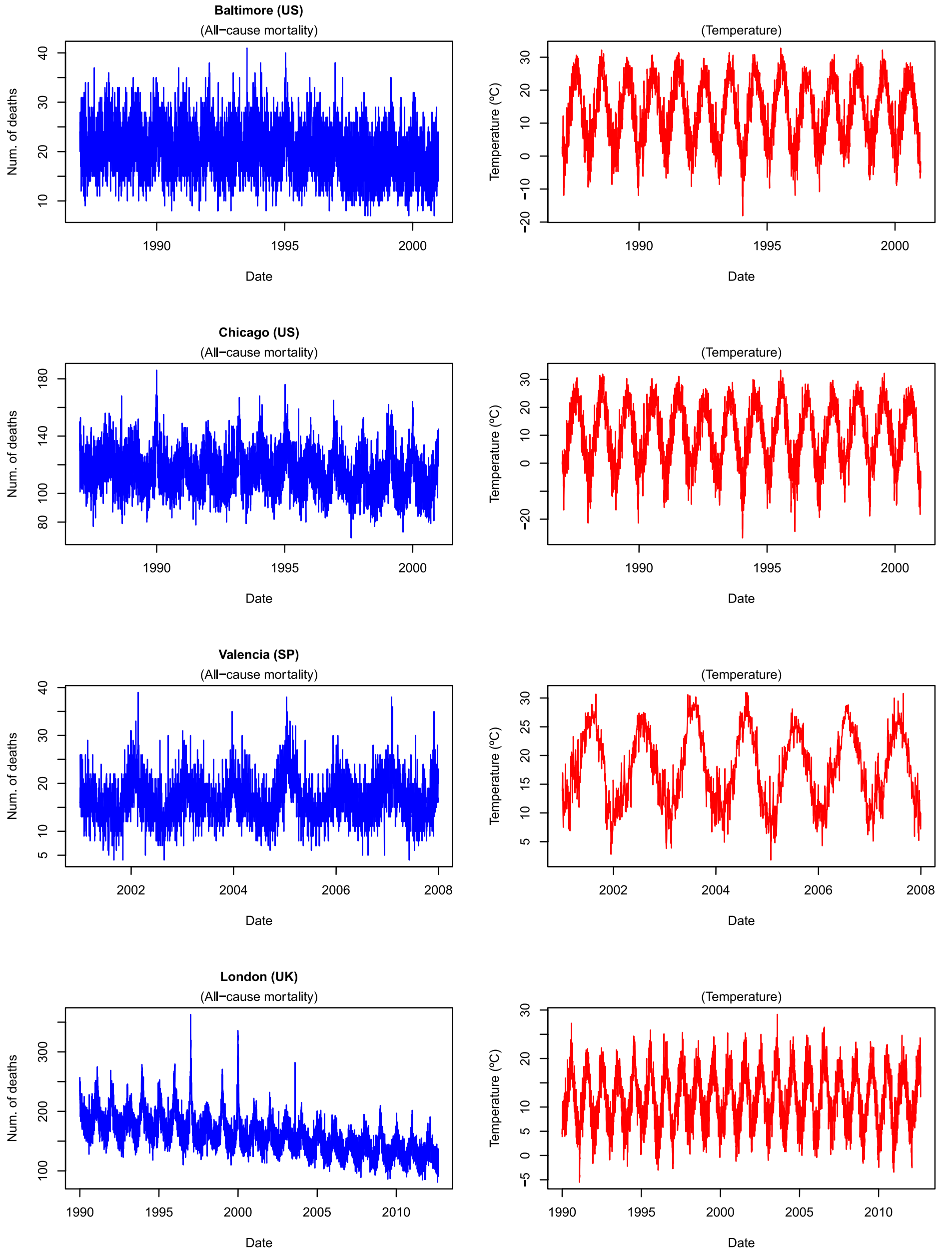


**Fig. 1.** Time-series for daily average temperature and all-cause mortality counts.

Fig. 2 shows the overall cumulative city-specific temperature-mortality associations across all lags. The relative risk of mortality is significantly higher for hot temperatures in Baltimore, whereas in Valencia, the opposite trend is observed, with a significantly higher risk for cold temperatures. In London, the mortality risks due to cold and heat appear to be of comparable magnitude, with the risk from cold being similar to that in Valencia and the risk from heat to

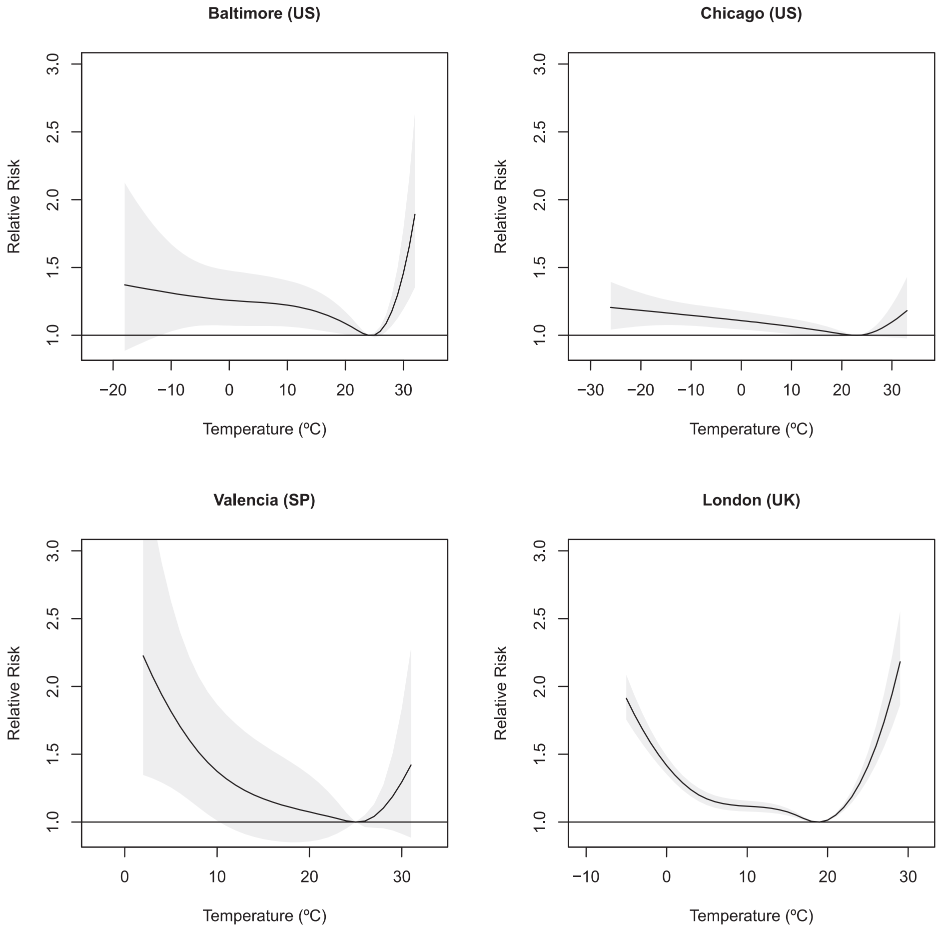


**Fig. 2.** Overall cumulative temperature-mortality association across all lags.

that in Chicago. Finally, in Baltimore, although the risks from cold and heat are similar, both are much lower compared to those in the other cities.

In the second stage, location-specific exposure-response curves were pooled using a random-effects multilevel meta-analytical model [15]. We used the geographical region (e.g., two cities in America and two cities in Europe) as a meta-predictor to account for variations in risk across locations.

Fig. 3 shows the predicted temperature-mortality association by geographical region, with America pooling exposure-response curves from Chicago and Baltimore, and Europe pooling those from Valencia and London. The pooled curve for the European cities shows a U-shape, indicating a higher mortality risk for cold than hot temperatures. For the American cities, the pooled curve shows a J-shape,with a slightly higher risk for heat compared to cold. However, the mortality risk for cold temperatures is much larger in Europe than in America, while the risk for heat is also slightly higher in European cities.

This case study demonstrated the applicability of the compiled time-series datasets for basic time-series regression analysis to estimate the temperature-mortality association. The availability of time-series data from multiple geographical locations also enabled the exploration of geographical differences by pooling location-specific associations within geographical regions.

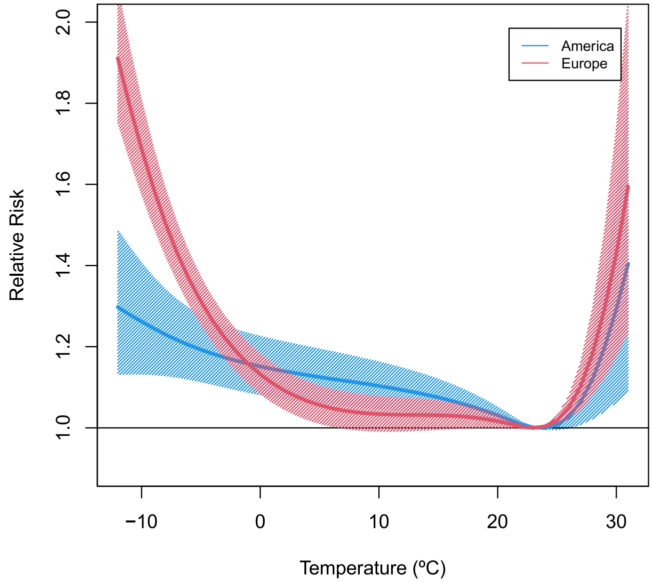


**Fig. 3.** Predicted temperature-mortality association by geographical region.

However, nowadays, reanalysis temperature data, such as ERA5, are commonly used in environmental epidemiology studies [16]. Therefore, collecting ground observation ambient temperature data from monitoring stations is not always necessary when daily mortality data is available. Additionally, the compiled time-series datasets can serve as valuable resources for teaching and as exemplars for research, not only for time-series regression studies but also for time-stratified case-crossover studies using aggregated data in environmental epidemiology [17].

## Limitations

Not applicable.

## Ethics Statement

The authors have read and follow the ethical requirements for publication in Data in Brief and confirming that the current work does not involve human subjects, animal experiments, or any data collected from social media platforms.

## Data Availability

tsdatasets (Reference data) (Mendeley Data).

## CRediT Author Statement

**Aurelio Tobias:** Conceptualization, Writing – original draft, Software, Methodology; **Chris Fook Sheng Ng:** Writing – review & editing; **Yoonhee Kim:** Writing – review & editing; **Masahiro Hashizume:** Writing – review & editing; **Lina Madaniyazi:** Conceptualization, Writing – review & editing.

## Acknowledgements

This compilation of time-series datasets has been used at the Asia Summer School on time-series regression for public health (https://asiasummerschool.weebly.com) for examples, exercises and working group activities. Aurelio Tobias was supported by the JSPS KAKENHI Grant Number 24K13527.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

[1] K. Bhaskaran, A. Gasparrini, S. Hajat, L. Smeeth, B. Armstrong, Time series regression studies in environmental epidemiology, Int. J. Epidemiol. 42 (4) (2013) 1187–1195.
[2] A. Gasparrini, B. Armstrong, M.G. Kenward, Distributed lag non-linear models, Stat. Med. 29 (21) (2010) 2224–2234.
[3] M.B. Mathur, M.P. Fox, Toward open and reproducible epidemiology, Am. J. Epidemiol. 192 (4) (2023) 658–664.
[4] S.M. Powers, S.E. Hampton, Open science, reproducibility, and transparency in ecology, Ecol. Appl. 29 (1) (2019) e01822.
[5] R. Peng, A. McDermott, tsModel: Time Series Modeling for Air Pollution and Health. R package version 0.6-1, 2022 https://CRAN.R-project.org/package=tsModel.
[6] A. Gasparrini, Distributed lag linear and non-linear models in R: the Package dlnm, J. Stat. Soft. 43 (8) (2011) 1–20.
[7] C. Iniguez, F. Ballester, A. Tobias, Data supporting the short-term health effects of temperature and air pollution in Valencia, Spain. Data Brief. 44 (2022) 108518.
[8] A. Gasparrini, Y. Guo, M. Hashizume, E. Lavigne, A. Tobias, A. Zanobetti, et al., Changes in susceptibility to heat during the summer: a multicountry analysis, Am. J. Epidemiol. 183 (11) (2016) 1027–1036.
[9] A.M. Vicedo-Cabrera, B. Forsberg, A. Tobias, A. Zanobetti, J. Schwartz, B. Armstrong, A. Gasparrini, Associations of inter- and intraday temperature change with mortality, Am. J. Epidemiol. 183 (4) (2016) 286–293.
[10] A.M. Vicedo-Cabrera, F. Sera, A. Gasparrini, Hands-on tutorial on a modeling framework for projections of climate change impacts on health, Epidemiol 30 (3) (2019) 321–329.
[11] T. Lytras, FluMoDL: Influenza-Attributable Mortality with Distributed-Lag Models, 2019. R package version 0.0.3, < https://CRAN.R-project.org/package=FluMoDL>
[12] T. Lytras, K. Pantavou K, E. Mouratidou, S. Tsiodras, Mortality attributable to seasonal influenza in Greece, 2013 to 2017: variation by type/subtype and age, and a possible harvesting effect, Euro. Surveill. 24 (14) (2019) 1800118.
[13] A. Gasparrini, Y. Guo, M. Hashizume, E. Lavigne, A. Zanobetti, J. Schwartz J, et al., Mortality risk attributable to high and low ambient temperature: a multicountry observational study, Lancet 386 (9991) (2015) 369–375.
[14] F. Sera, A. Gasparrini, Extended two-stage designs for environmental research, Environ. Health 21 (1) (2022) 41.
[15] F. Sera, B. Armstrong, M. Blangiardo, A. Gasparrini, An extended mixed-effects framework for meta-analysis, Stat. Med. 38 (29) (2019) 5429–5444.
[16] D. Roye, C. Iñiguez C, A. Tobias, Comparison of temperature-mortality associations using observed weather station and reanalysis data in 52 Spanish cities, Environ. Res. 183 (2020) 109237.
[17] A. Tobias, Y. Kim, L. Madaniyazi, Time-stratified case-crossover studies for aggregated data in environmental epidemiology: a tutorial, Int. J. Epidemiol. 53 (2) (2024) dyae020.