THE LANCET
Planetary Health

Supplementary appendix

This appendix formed part of the original submission and has been peer reviewed. We post it as supplied by the authors.

Supplement to: Gasparrini A, Masselot P, Scortichini M, et al. Small-area assessment of temperature-related mortality risks in England and Wales: a case time series analysis. *Lancet Planet Health* 2022; **6:** e557–64.

Supplementary Material of the article:

Small-area assessment of temperature-related mortality risks: a case time series analysis in England and Wales

Statistical annexe

Collection and definition of small-area characteristics

We collected information on a series of small-area variables defined for lower super output areas (LSOAs) and then aggregated them at local authority districts (LADs). The LSOAs are census-based statistical units with approximately 1,600 residents, while LADs are administrative subdivisions. These areas correspond to the definitions used in the 2011 census, with a total of 34,753 LSOAs nested in 348 LADs. The selected characteristics include the 15 small-area variables combined in the principal component analysis (PCA) to derive composite indicators of vulnerability to heat and cold (detailed in the next sections), and other attributes used for aggregating the results.

The data were gathered from multiple sources. Specifically:

- Average daily temperature and its annual range (difference between the maximum and minimum values across the study period), both in °C, were directly computed as summaries of the LSOA-specific daily temperature series across the study period. The series was derived as the average of daily minimum and maximum temperatures from the HadUK-Grid 1x1km gridded daily data from Met Office (version 1.0.3.0),[1,2] available at the CEDA web repository (http://archive.ceda.ac.uk/).
- Data on demographic variables (proportion of population above 65 in %, and population density as the number of persons per hectare) and proportion of houses with central heating (in %) were obtained by key statistics datasets (KS101EW, KS102EW, and KS403EW, respectively) of the 2011 census from NOMIS (https://www.nomisweb.co.uk/).
- Variables related to the index of multiple deprivation (IMD) were extracted from governmental databases for England (www.gov.uk/, for the year 2015) and Wales (https://gov.wales, for the year 2014). These variables represent IMD sub-domains corresponding to various aspects, specifically income, employment, education, health & disability, access to housing & services, and living environment. These were quantified as ranks of LSOAs (from the most to the least deprived for each sub-domain) computed separately for England and Wales.
- Topographical characteristics related to land use, available at different spatial resolutions as gridded products, were retrieved from NASA's earth observation satellites and Copernicus Land Monitoring Service (CLMS). Specifically, we collected measures of albedo, an indicator of direct or diffuse solar radiation reflected by Earth's surface (https://ladsweb.modaps.eosdis.nasa.gov/missions-and-measurements/products/MCD43A3), surface imperviousness density, a measure of the relative presence of artificial structures (https://land.copernicus.eu/pan-european/high-resolution-layers/imperviousness/statusmaps/2012), and enhanced vegetation index, an improved measure that quantifies the amount of vegetation in a defined area (https://ladsweb.modaps.eosdis.nasa.gov/missions-and-measurements/products/MOD13Q1). These indices are measured on a scale of 0-1 during 2011 and were assigned for each LSOA using area-weighted averages of the intersecting grid cells.
- The other infrastructural variable is the average age of buildings (in years). The original dataset was gathered from the Consumer Data Research Center and contains the number of residential dwellings by 10-year age groups from pre-1900 to 2015 at the LSOA level (https://data.cdrc.ac.uk/dataset/dwelling-ages-and-prices).
- The urban/rural classification for each LSOA at the 2011 census was collected from the Office of National Statistics (www.ons.gov.uk/), while the full IMD score was available from the

NOMIS database (see above). These two were used as stratification variables, using five rural/urban categories (major conurbations, minor conurbation, urban cities and towns, rural towns, rural villages) and IMD quintiles. The resident population in each LSOA stratified in age groups for the year 2011 was collected from census data in NOMIS (see above), and used as the denominator for the computation of the standardised rates.

Extended two-stage design and statistical models

We applied an analytical framework based on an extended two-stage design. The first-stage models were performed separately in four age groups (0-64, 65-74, 75-84, and 85 and older) for each of the 348 LADs, and the results were then pooled in a second-stage meta-regression.

In the first stage, we applied a novel design, called case time series, to model data collected at small area level, specifically series of temperature and mortality defined for all the LSOAs within each LAD.[3] The data were fitted using a conditional Poisson regression,[4] where the expected counts of deaths $\mathrm{E}(y_{jait})$ at day t in LSOA i were modelled separately for each age group a within LAD j as:

$$g[\mathrm{E}(y_{jait})] = \xi_{jaik} + f(x_{jit}, \ell; \boldsymbol{\theta}_{ja}) + s(t; \boldsymbol{\gamma}_{ja}) + I(w_t; \boldsymbol{\varphi}_{ja})$$

where g is a logarithm link function, $s(t)$ is a natural spline function of time with 1 degree of freedom (df) per year and a vector of coefficients $\boldsymbol{\gamma}_{ja}$ to model age and LAD-specific long-term trends common to all LSOAs within LAD j, and $I(w_t)$ are indicators for the day of the week w with coefficients $\boldsymbol{\varphi}_{ja}$. We used a distributed lag non-linear model (DLNM) to represent the non-linear and lagged dependency between mortality and temperature x_{jit} (as an exposure common to all age groups). The DLNM defines a bi-dimensional function $f(x_{jit}, \ell)$ with a vector of m coefficients $\boldsymbol{\theta}_{ja}$ that represents an exposure-lag-response risk surface along with the spaces of predictor and lag.[5] Specifically, this is defined using a cross-basis parameterization of spline functions, using quadratic B-splines with three internal knots placed at the 10th, 75th, and 90th percentiles of LAD-specific temperature distributions for the exposure-response, and natural cubic splines with three internal knots placed at equally spaced values in the log scale of the lag window 0-21 (plus intercept) for the lag-response. This model definition was selected in previous temperature-mortality studies.[6] Finally, the intercepts ξ_{jaik} are defined for each stratum k corresponding to month-in-year subsets of the study period, and they capture LSOA-specific deviations in risk above LAD-specific trends modelled by $s(t)$. These high number of intercepts represent nuisance parameters which are conditioned out and not estimated in the conditional Poisson regression. This flexible case time series model allows the application of sophisticated time series techniques such as smoothing and DLNMs at small-area level, using multiple series of high-resolution data defined at LSOA and enabling a fine control for temporal trends.

In the second stage, the age and LAD-specific estimated parameters $\widehat{\boldsymbol{\theta}}_{ja}$ that define complex bi-dimensional temperature-mortality associations are first reduced to the related parameters $\widehat{\boldsymbol{\theta}}^{*}_{ja}$ of the overall cumulative exposure-response function representing the net effects cumulated across lags.[7] These are then pooled using a multivariate mixed-effects meta-regression,[8] algebraically formulated as:

$$\widehat{\boldsymbol{\theta}}^{*}_{ja} = \boldsymbol{X}_{ja}\boldsymbol{\beta} + \boldsymbol{Z}_{ja}\boldsymbol{b}_{ja} + \boldsymbol{\varepsilon}_{ja}$$

The model allows multivariate outcomes and repeated measurements for different age groups within the same LAD. A detailed description of the modelling framework and algebra is available elsewhere.[8] Here, the first term on the right-hand side represents fixed-effects average associations defined in terms of meta-predictors in the design matrix $\boldsymbol{X}_{ja}$ and coefficients $\boldsymbol{\beta}$. The second term defines potential random-effects deviations across LADs with the related design matrix $\boldsymbol{Z}_{ja}$ and coefficients $\boldsymbol{b}_{ja}$, while the last term $\boldsymbol{\varepsilon}_{ja}$ represents age/LAD-level sampling errors dependent on the (co)variance matrix of $\widehat{\boldsymbol{\theta}}^{*}_{ja}$. The design matrix $\boldsymbol{X}_{ja} = \boldsymbol{x}^{T}_{ja} \otimes I_m$ is a multivariate expansion of fixed-effects meta-regressors $\boldsymbol{x}_{ja} = [\boldsymbol{x}^{T}_{j}, \boldsymbol{x}^{T}_{a}]^{T}$. Specifically, $\boldsymbol{x}_j$ represents the LAD-level values of the first p principal components derived from the 15 small-area variables $\boldsymbol{w}_j$ (see below for details). Here, $\boldsymbol{x}_j$ are included as composite vulnerability indicators to explain variations in temperature-mortality associations related to differential geographical patterns in risk. The other meta-regressors $\boldsymbol{x}_a$

correspond to age group indicators to specify age-specific exposure-responses. As the model with the age and PCA terms showed negligible residual heterogeneity (Cochran Q test: *p*-value = 0·49; I^2 = 0·1%), we simplified the analysis using a fixed-effects meta-regression, thus omitting the term $\boldsymbol{Z}_{ja}\boldsymbol{b}_{ja}$. This model was then used to downscale the estimated risk by predicting the parameters of the age-specific overall cumulative exposure-response relationship for each LSOA based on the values of the meta-predictors (PCA indicators) computed at this geographical level.

Analysis of vulnerability through PCA indicators

The idea underpinning the use of PCA to derive indicators of vulnerability is the possibility to reduce a potentially long list of correlated predictors in few uncorrelated components.[9] This decreases the dimensionality of the second-stage model and allows combining multiple modification effects from the small-area characteristics in composite vulnerability indicators. Starting from the original vector of 15 small-area variables $\boldsymbol{w}_j$, the PCA components $\boldsymbol{x}_j$ can be computed from a series of orthogonal rotations defined algebraically as:

$$\boldsymbol{x}_j = \boldsymbol{R}^T\boldsymbol{w}_j$$

Here, $\boldsymbol{R}$ is the 15×15 matrix of loadings defining the rotations, chosen such that all indicators in $\boldsymbol{x}_j$ are uncorrelated and their variance is maximal.[9] The first $p = 3$ components explained 70·4% of the variability and were selected for inclusion as meta-predictors. These indicators can be interpreted as unobservable traits that explain the observed correlation among indicators and combine their effects.

Computation of excess mortality and related uncertainty

The exposure-response functions predicted at each LSOA and age group were finally used to compute the excess mortality associated with non-optimal temperature. The methodology for quantifying the excess number of deaths and the related fraction from complex exposure-lag-response relationships, also separating the contributions of heat and cold, is illustrated in detail elsewhere.[10] This approach has been applied in several multi-location studies.[6] The excess number of deaths for each age group was finally divided by the number of residents in that group to compute age-specific rates, then standardised using as reference the Standard European Population for 2013. The standardisation was applied using as numerators and denominators the quantities aggregated at different geographical levels (LSOAs, LADs, regions) and categories (index of multiple deprivation and rural/urban classification). For LSOA, age-specific rates of total mortality at LAD level were used to avoid issues with small counts of mortality and resident population, and then multiplied by the estimated mortality fraction.

Uncertainty of estimated excess deaths and standardised rates was computed by extending a previously proposed method to account for the dependencies generated from the use of the same model to predict LAD-specific and downscaled LSOA-specific associations. If unaccounted for, these can lead to a severe underestimation of the uncertainty. In the original method,[6,10] the multivariate meta-regression model was used to derive best linear unbiased predictions (BLUPs) of the parameters of the exposure-response function, together with related (co)variance matrix, in each location j. These quantities were used to sample occurrences of parameters in Monte Carlo iterations, independently for each location. The simulated distribution of excess mortality using the set of iterations was then used to define empirical confidence intervals (eCI). Corresponding values for higher geographical levels were derived by aggregating the results from each iteration, summing, and finally computing point estimates and eCI.

In order to account for the stronger dependencies across locations in this small-area analytical setting, we instead performed 500 Monte Carlo iterations by sampling directly from the estimated coefficients $\widehat{\boldsymbol{\beta}}$ and associated (co)variance matrix of the multivariate repeated-measure fixed-effects meta-regression model. Therefore, for each iteration, the parameters of the LAD-specific and downscaled LSOA-specific temperature-mortality associations are predicted using the same simulated model, and not independently. Eventually, when aggregating at higher geographical

levels, their dependency is appropriately accounted for in order to retrieve the correct empirical distribution and eCI.

References

1. Hollis D, McCarthy M, Kendon M, Legg T, Simpson I. HadUK-Grid—A new UK dataset of gridded climate observations. *Geoscience Data Journal* 2019; **6**(2): 151-9.
2. Met Office, Hollis D, McCarthy M, Kendon M, Legg T, Simpson I. HadUK-Grid Gridded Climate Observations on a 1km grid over the UK, v1.0.2.1 (1862-2019): Centre for Environmental Data Analysis, 2020.
3. Gasparrini A. The case time series design. *Epidemiology* 2021: DOI: 10.1097/EDE.0000000000001410.
4. Armstrong BG, Gasparrini A, Tobias A. Conditional Poisson models: a flexible alternative to conditional logistic case cross-over analysis. *BMC Medical Research Methodology* 2014; **14**(1): 122.
5. Gasparrini A. Modeling exposure-lag-response associations with distributed lag non-linear models. *Statistics in Medicine* 2014; **33**(5): 881-99.
6. Gasparrini A, Guo Y, Hashizume M, et al. Mortality risk attributable to high and low ambient temperature: a multicountry observational study. *The Lancet* 2015; **386**(9991): 369-75.
7. Gasparrini A, Armstrong B. Reducing and meta-analyzing estimates from distributed lag non-linear models. *BMC Medical Research Methodology* 2013; **13**(1): 1.
8. Sera F, Armstrong B, Blangiardo M, Gasparrini A. An extended mixed-effects framework for meta-analysis. *Statistics in Medicine* 2019; **38**(29): 5429-44.
9. Joliffe IT. Principal Component Analysis. 2nd ed. New York: Springer; 2002.
10. Gasparrini A, Leone M. Attributable risk from distributed lag models. *BMC Medical Research Methodology* 2014; **14**(1): 55.

Supplementary Tables

Table S1. Distribution of relative risks (RR), excess mortality fractions, minimum mortality temperatures and percentiles (MMT and MMP), and standardised excess mortality rates across lower super output areas (LSOAs) in England & Wales (2000-2019) by age groups. The mean values correspond to population-weighted averages.

Effect	Age group	Mean	Min	25th	Median	75th	Max
RR (1st and 99th vs MMT)							
Cold	0-64	1·277	1·144	1·243	1·271	1·307	1·523
	65-74	1·388	1·223	1·344	1·380	1·426	1·696
	75-84	1·461	1·306	1·422	1·455	1·496	1·739
	85plus	1·561	1·404	1·524	1·555	1·594	1·850
Heat	0-64	1·057	1·000	1·020	1·045	1·081	1·363
	65-74	1·041	1·000	1·004	1·028	1·062	1·340
	75-84	1·054	1·000	1·018	1·042	1·078	1·357
	85plus	1·098	1·000	1·059	1·088	1·128	1·423
Excess mortality fraction (%)							
Cold	0-64	6·43	2·47	5·38	6·29	7·31	17·42
	65-74	8·52	0·47	7·28	8·30	9·47	25·84
	75-84	11·51	1·99	10·45	11·43	12·46	22·71
	85plus	14·31	2·36	13·34	14·33	15·32	28·35
	all	10·88	4·23	9·55	10·88	12·18	20·21
Heat	0-64	0·16	-0·13	0·04	0·11	0·21	2·55
	65-74	0·11	-0·25	0·01	0·06	0·15	6·12
	75-84	0·14	-0·14	0·04	0·10	0·19	5·70
	85plus	0·24	-0·04	0·12	0·19	0·30	7·69
	all	0·17	-0·07	0·08	0·14	0·23	1·85
MMP (%) and MMT (°C)							
MMP	0-64	95	92	94	95	96	99
	65-74	95	86	93	94	97	99
	75-84	95	92	94	95	96	99
	85plus	95	92	94	94	95	99
MMT	0-64	18·9	15·9	18·3	18·9	19·5	22·4
	65-74	19·0	15·2	18·3	19·0	19·7	22·6
	75-84	19·0	16·0	18·4	19·0	19·6	22·4
	85plus	18·6	14·9	18·1	18·6	19·2	21·7
Standardised excess mortality rate (x 100,000 person-years)							
Cold	-	120·20	1·34	109·92	119·69	130·58	242·03
Heat	-	1·80	-0·74	0·83	1·51	2·46	13·12

Supplementary Figures

Figure S1. Flowchart describing the modelling framework. The different shapes represent input/output data (parallelograms) and processes/models (rectangles), with the original input data (blue), derived input data (grey), analytical steps (orange), and output estimates (green).

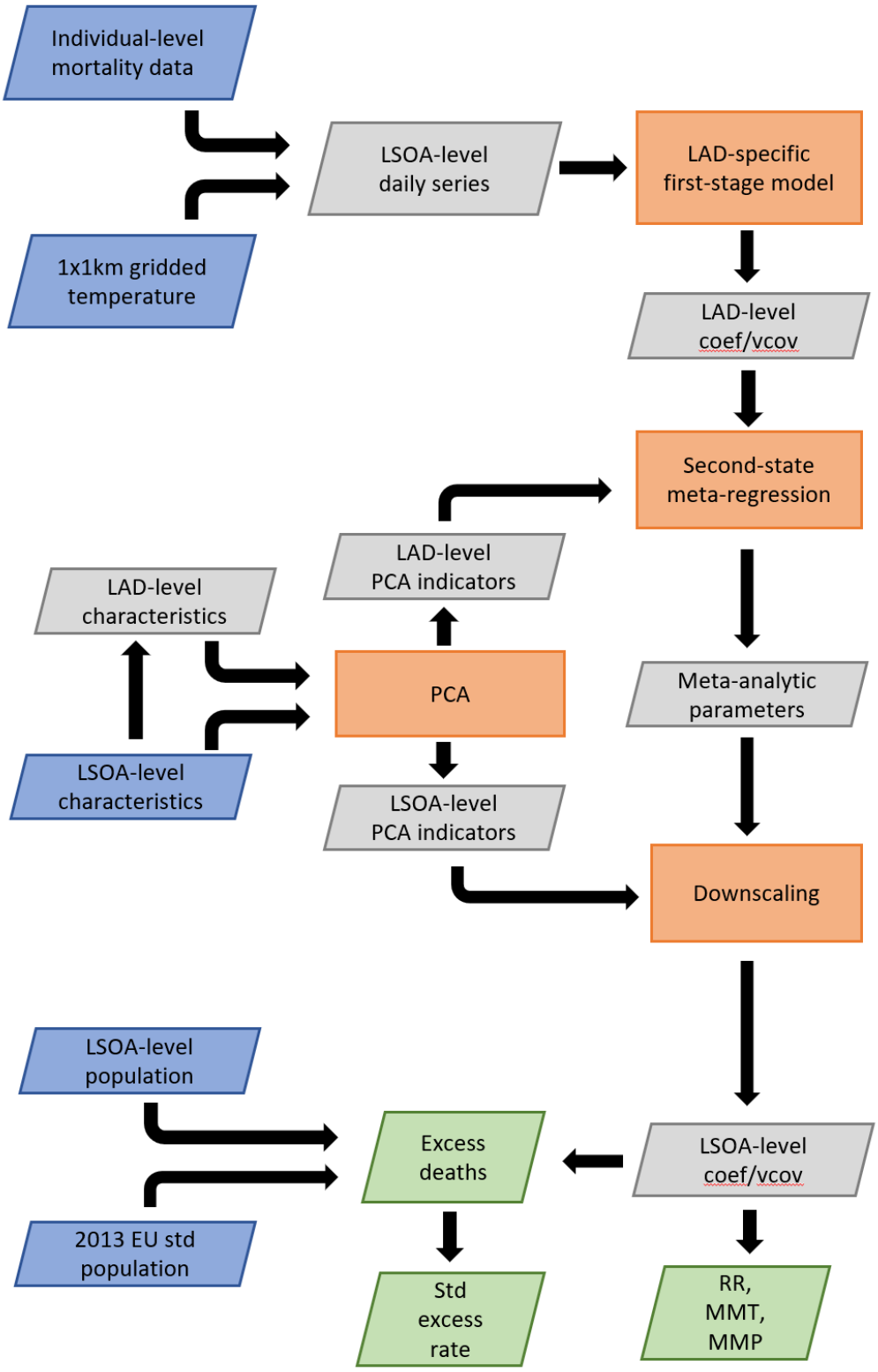

Figure S2. Correlation plots of the 15 small-area variables at local authority district (LAD, left panel) and lower super output area (LSOA, right panel) levels.

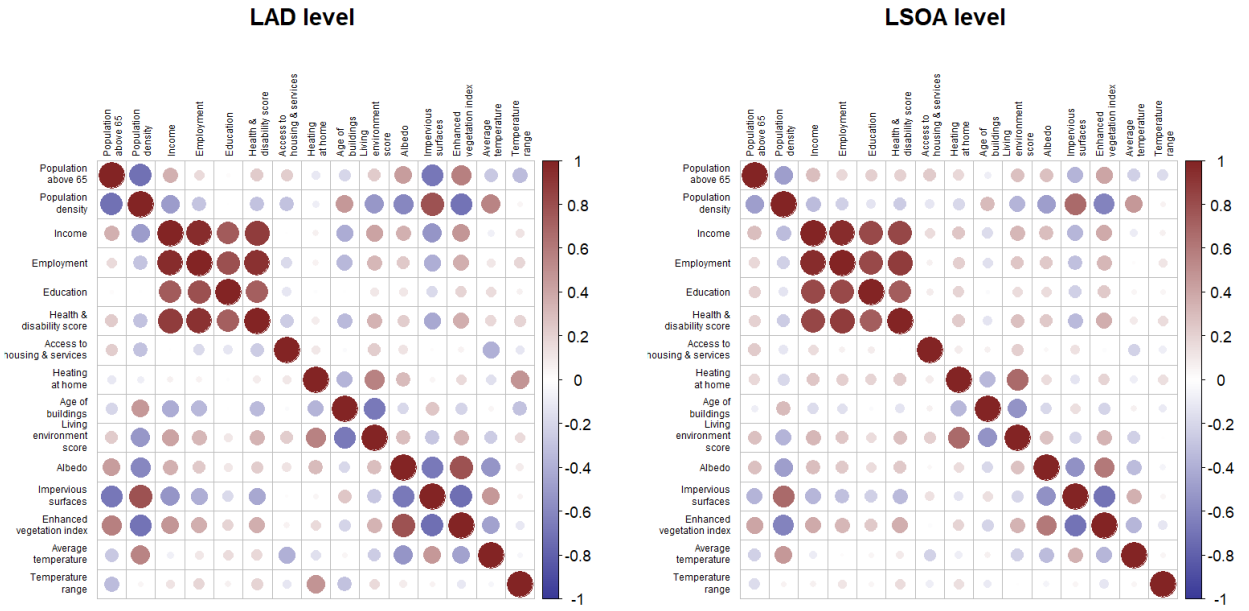

Figure S3. Screeplot (left panel) and cumulative variance plot (right panel) of the principal component analysis (PCA) of the 15 small-area variables at the local authority district (LAD) level.

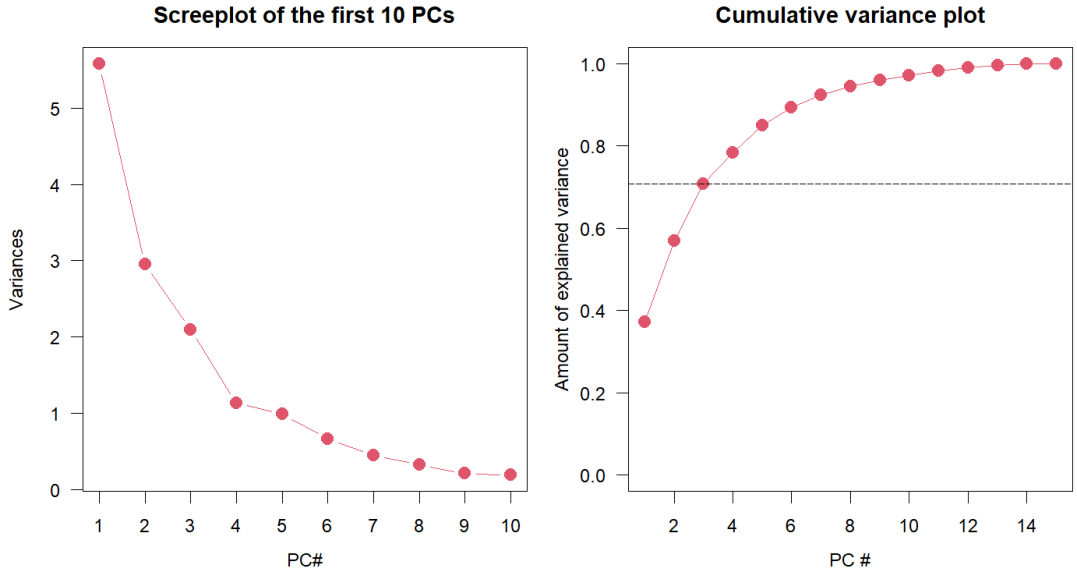

Figure S4. Heatmap representing the correlation between the input and output of the first six principal components (*y*-axis) derived from the original 15 small-area variables (*x*-axis) through a principal component analysis (PCA). The first three (highlighted) components capture most of the correlations (PCA scores) with the original variables and were used as vulnerability indicators in the meta-regression.

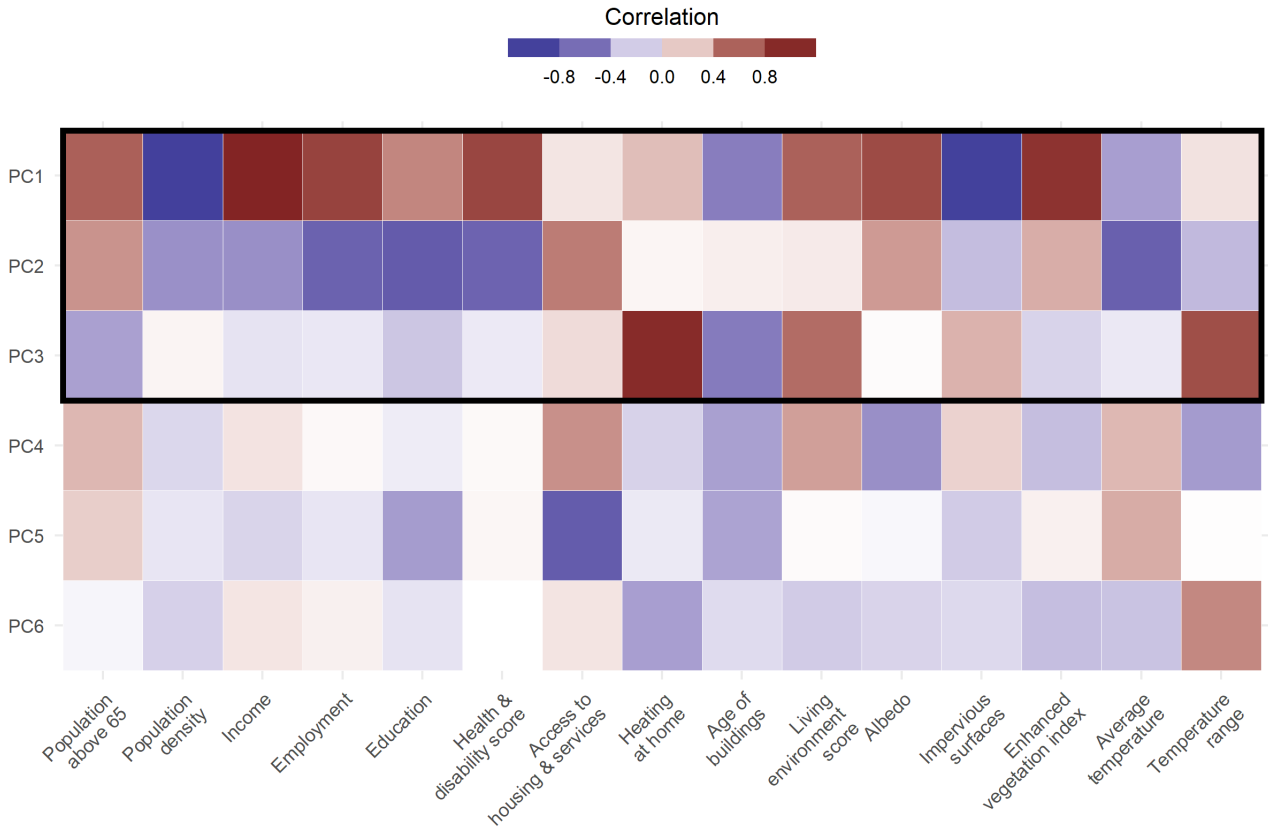

Figure S5. Geographical subdivisions corresponding to the nine administrative regions of England and the country of Wales.

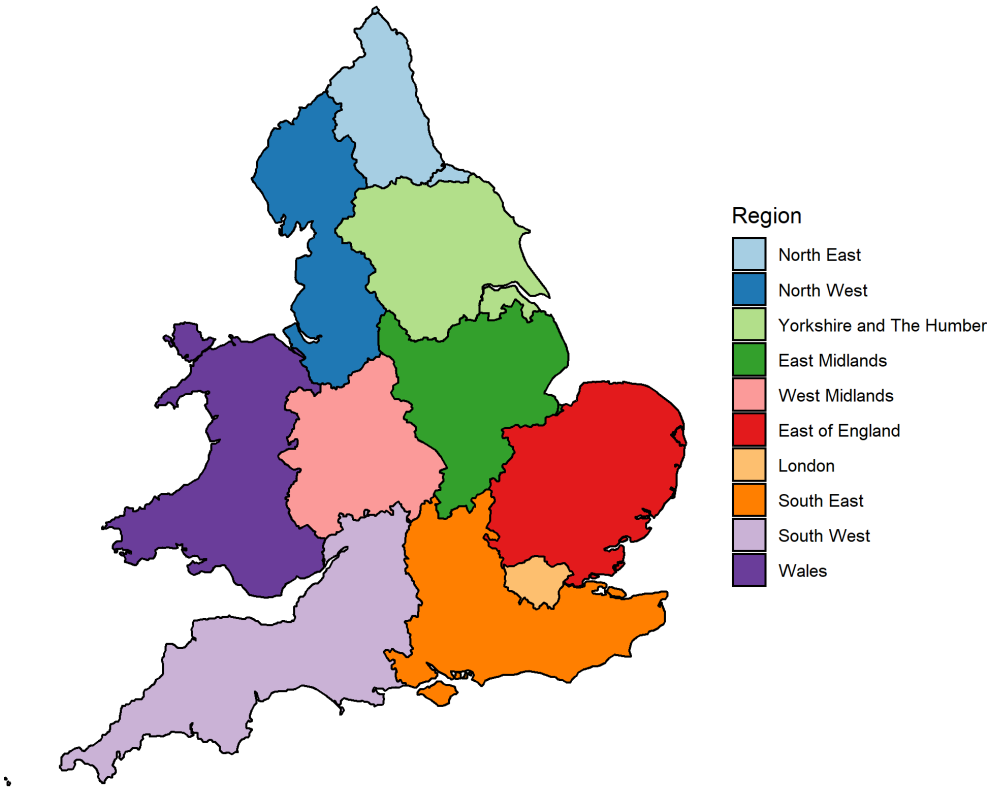

Figure S6. Excess mortality fraction (with 95% empirical confidence intervals, eCI, in %) attributable to heat (top panel) and cold (bottom panel) in England & Wales, stratified by region, quintiles of the index of multiple deprivation (IMD), and urban/rural classification.

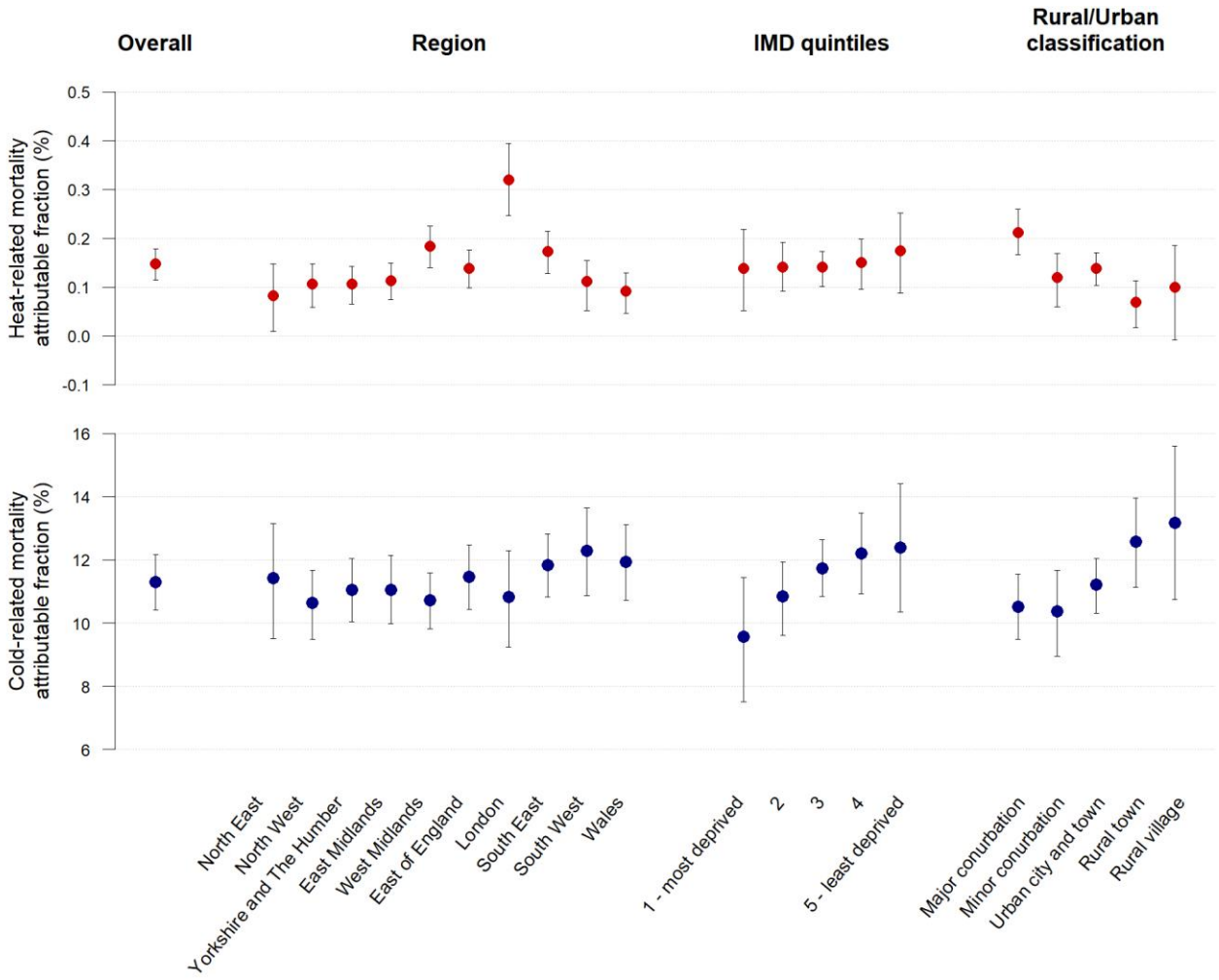

Figure S7. Geographical distribution of the 15 small-area variables across the 34,753 lower super output areas (LSOAs) of England & Wales (see the first section of this document for definitions and units).

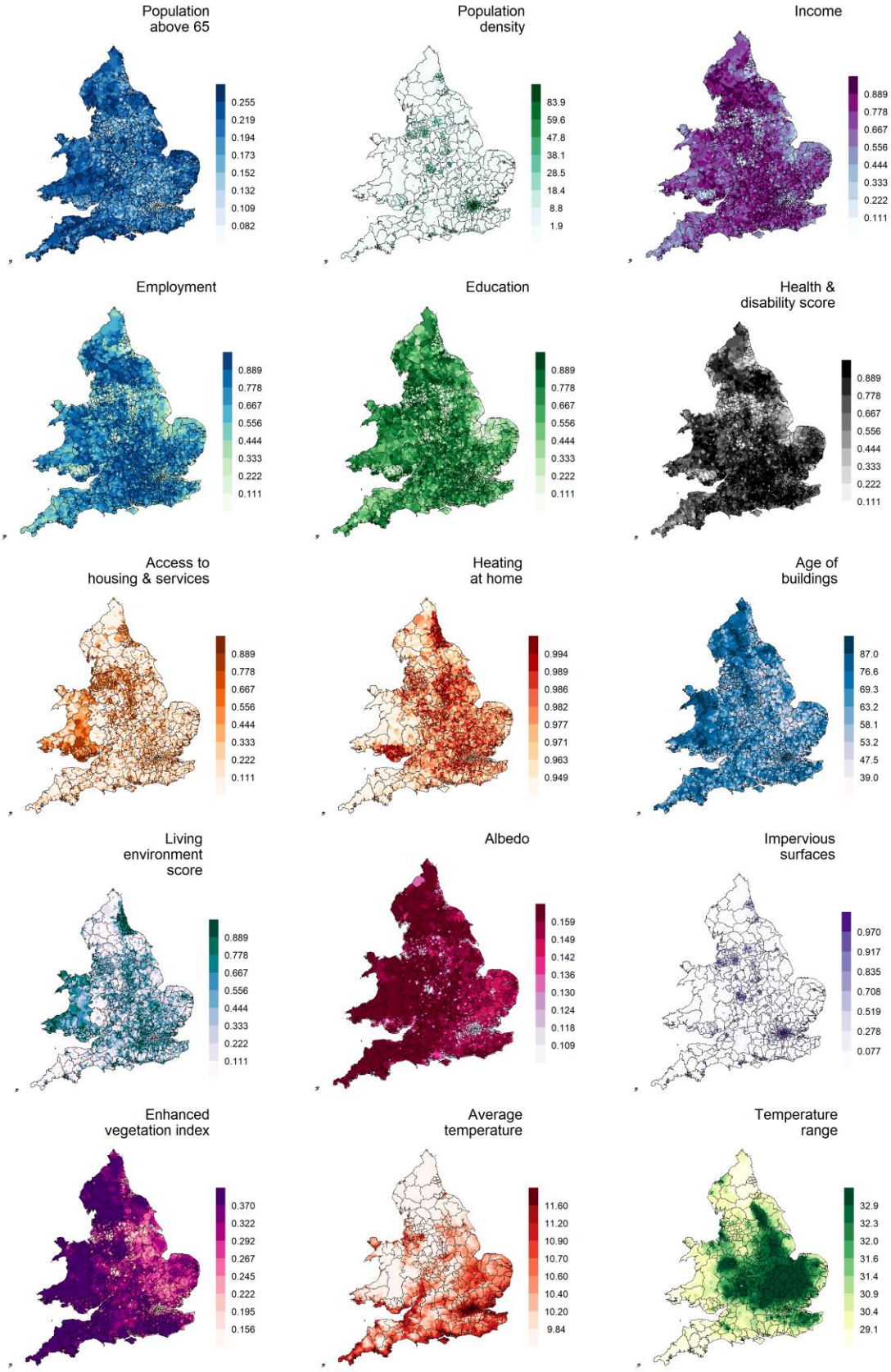

Figure S8. Geographical distribution of the first three principal components used as composite vulnerability indicators across the 34,753 lower super output areas (LSOAs) of England & Wales.

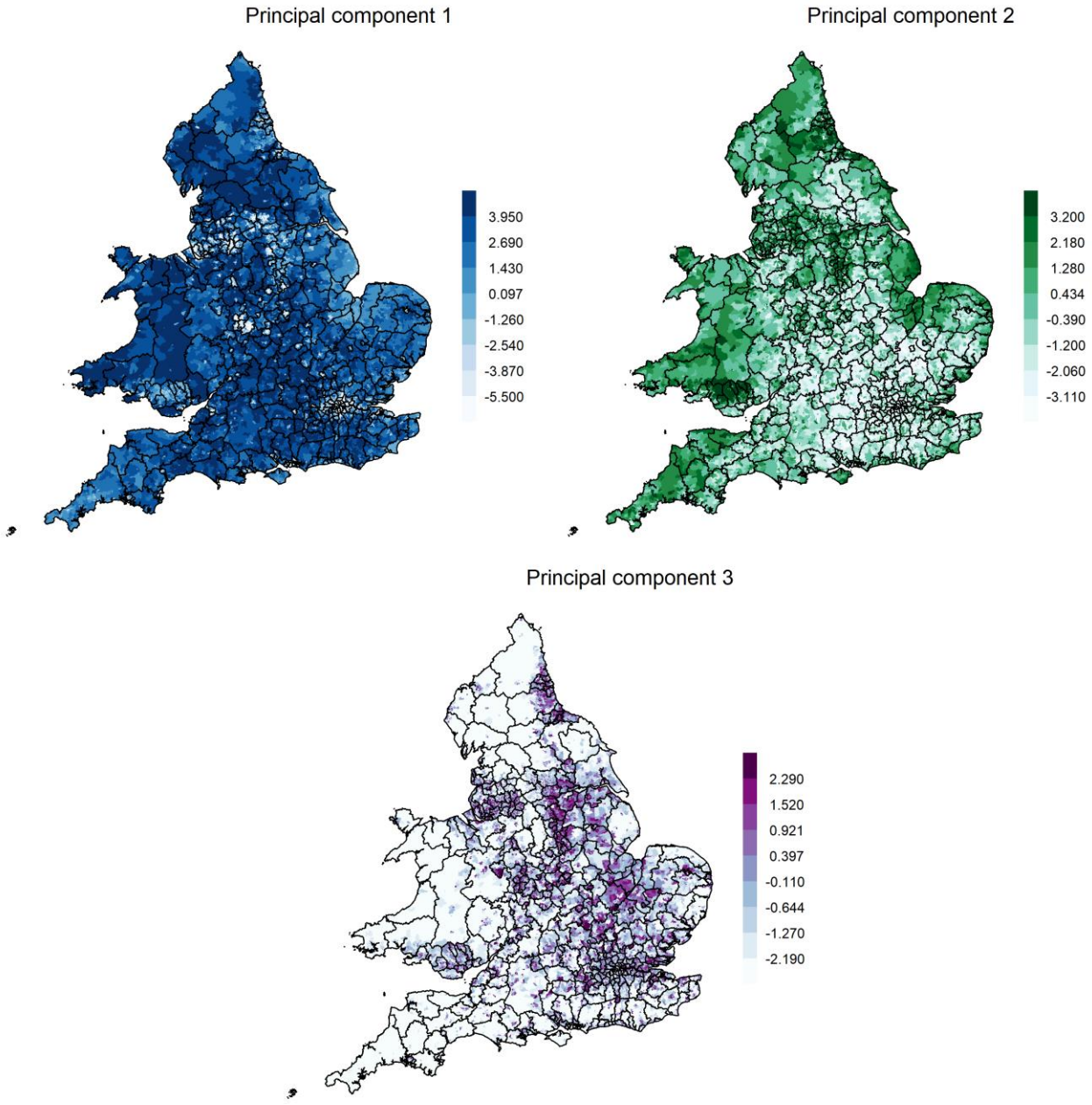

Figure S9. Geographical distribution of the relative risk (RR) associated with heat (99th temperature percentile vs MMT, left panels) and cold (1st vs MMT, right panel) by age group across lower super output areas (LSOAs) of England & Wales.

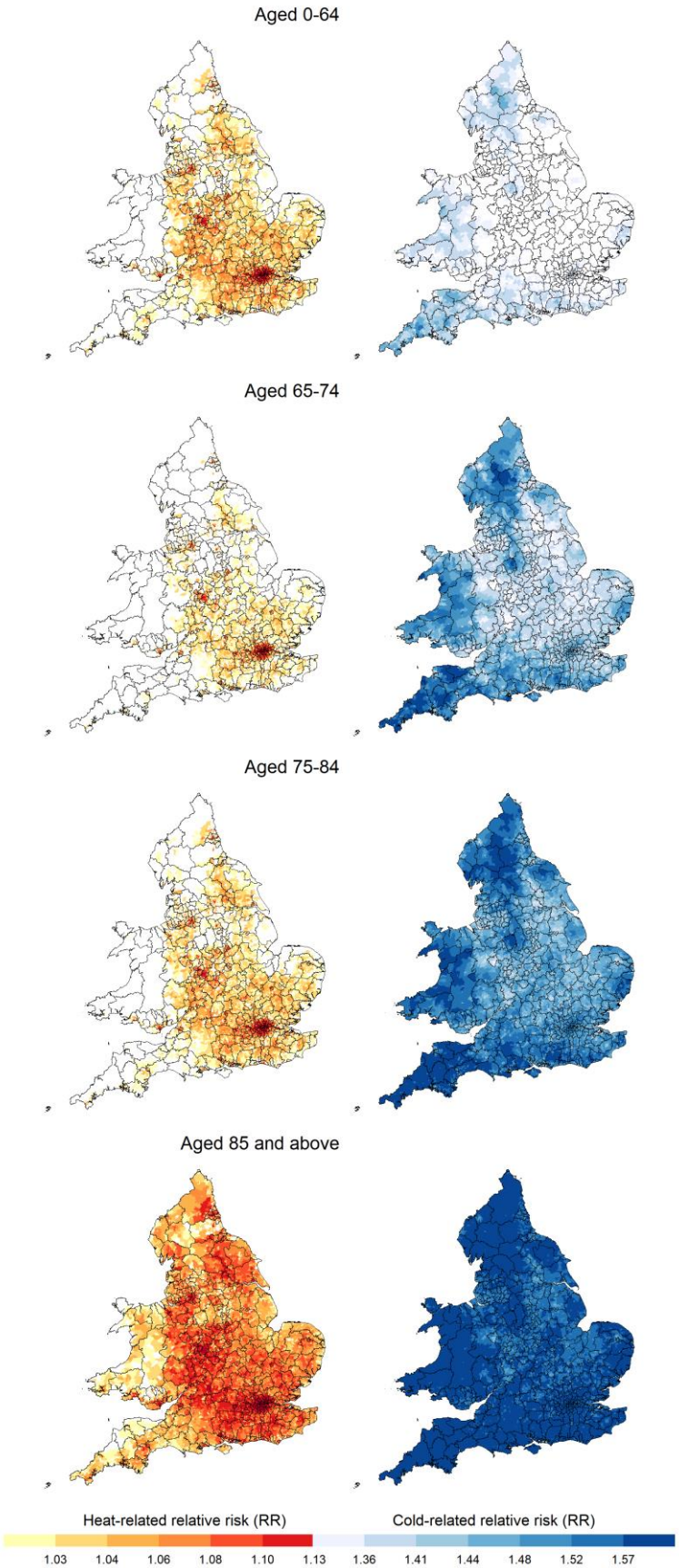

Figure S10. Geographical distribution of the minimum mortality temperature percentile (MMP, in %) across lower super output areas (LSOAs). The maps illustrate the distribution across the whole spatial domain of England & Wales (left panels), and then with a focus on Greater London (mid panel) and the local authority district (LAD) corresponding to the Borough of Camden (right panel).

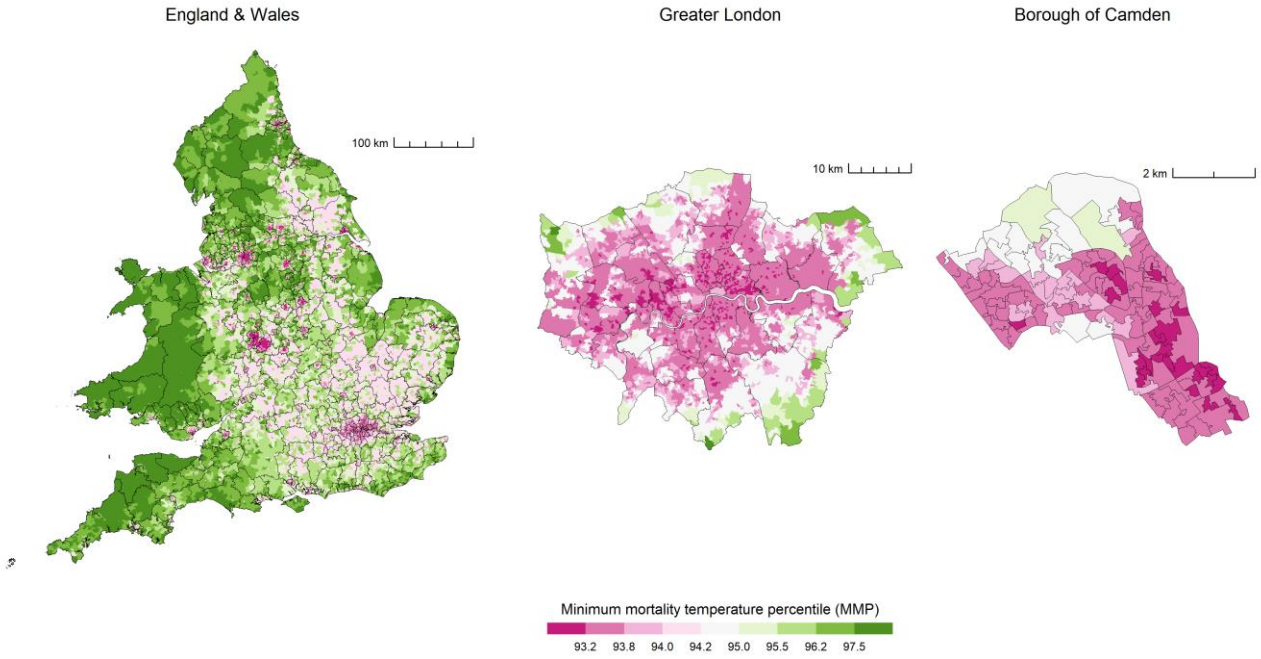
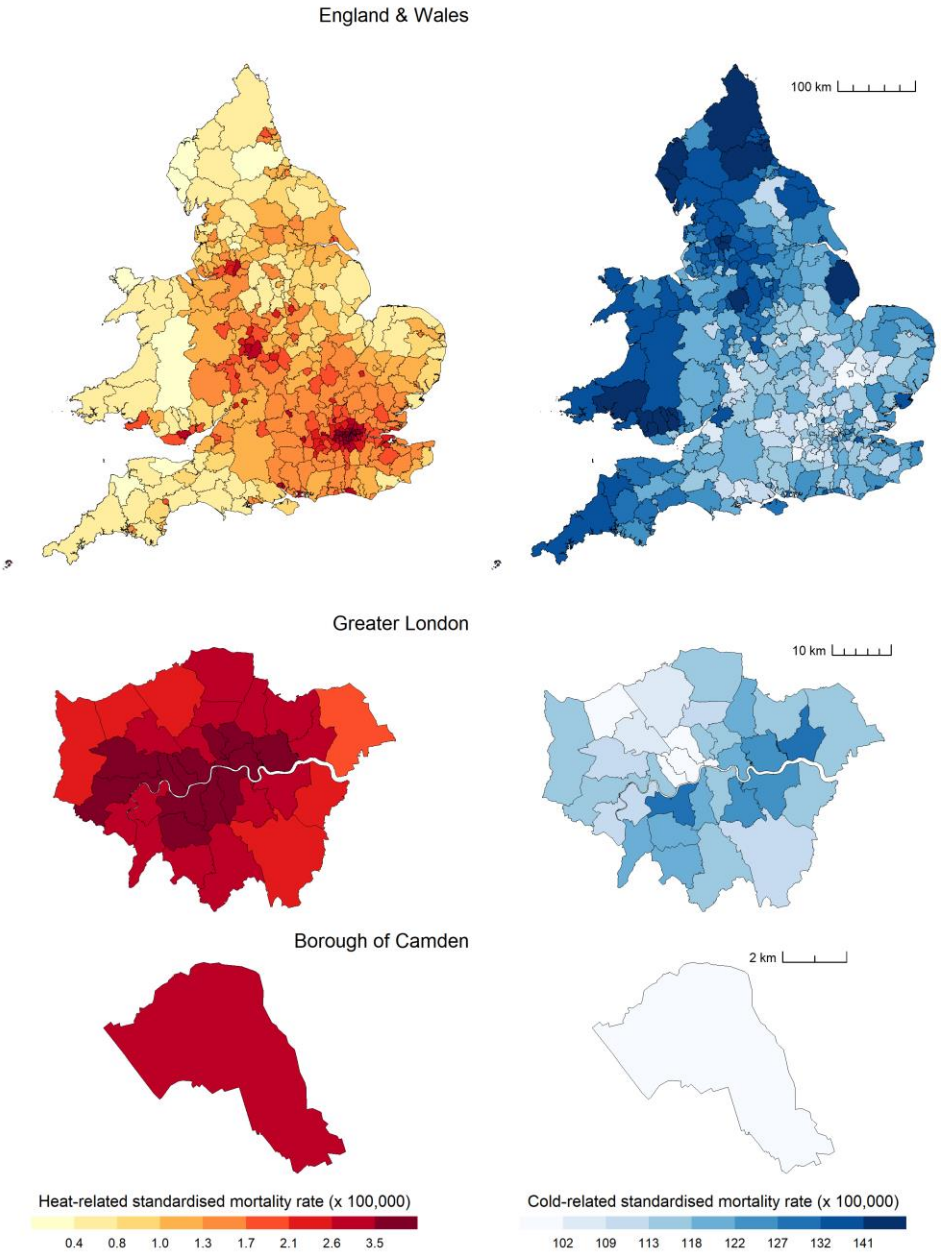
Figure S11. Geographical distribution of the standardised excess mortality rate (x 100,000 person-years) for heat and cold (left and right-column panels, respectively) across local authority districts (LADs). The maps illustrate the distribution across the whole spatial domain of England & Wales (top-row panels), and then with a focus on Greater London (mid-row panels) and the local authority district (LAD) corresponding to the Borough of Camden (bottom-row panels).