The Innovation, Volume *2*

Supplemental Information

Ambient Temperature and Years of Life Lost: A National Study in China

Tao Liu, Chunliang Zhou, Haoming Zhang, Biao Huang, Yanjun Xu, Lifeng Lin, Lijun Wang, Ruying Hu, Zhulin Hou, Yize Xiao, Junhua Li, Xiaojun Xu, Donghui Jin, Mingfang Qin, Qinglong Zhao, Weiwei Gong, Peng Yin, Yiqing Xu, Jianxiong Hu, Jianpeng Xiao, Weilin Zeng, Xing Li, Siqi Chen, Lingchuan Guo, Zuhua Rong, Yonghui Zhang, Cunrui Huang, Yaodong Du, Yuming Guo, Shannon Rutherford, Min Yu, Maigeng Zhou, and Wenjun Ma

Supplemental Information

1.Supplementary text

1.1 China's Disease Surveillance Points system and study location selection

1.2 Interpolation of daily meteorological factors

1.3 Daily PM_{10} estimation using a land-use-regression (LUR) model

1.4 Estimation of the associations of temperatures and TVs with YLL rates

1.5 Calculations of the YLL per death due to temperatures

1.6 Calculations of the annual mean YLL due to temperatures in China

2.Supplementary tables

Table S1. General characteristics of included 364 locations in China

Table S2. YLL per death due to different components of temperatures across 364 locations in China

3.Supplementary figures

Figure S1. The geographic distribution of 364 locations in China

Figure S2. Scatter plot of 10-fold cross-validation of interpolated meteorological factors in 698 weather stations in China

Figure S3. The process of daily PM_{10} concentration estimation

Figure S4. Performance and validation of estimating the daily PM_{10} concentration in China

Figure S5. The distribution of daily average YLL rate (/100,000 population) in 364 locations in China

Figure S6. Overall lag structure in effects of extreme cold temperature on daily YLL rates in all 364 locations and in different subgroups in China

Figure S7. Overall lag structure in effects of extreme hot temperature on daily YLL rates in all 364 locations and in different subgroups in China

Figure S8. Univariable meta-regression results of the modification effects of city level characteristics on the YLL per death caused by temperatures in 364 locations in China

Figure S9. Sensitivity analyses on the impacts of lag days and df (/year) on the nationwide exposure-response relationship between temperature and YLL rate in China

1.Supplementary text

1.1 China's Disease Surveillance Points system and study location selection

The China's Disease Surveillance Points System (DSPS) is administrated by the Chinese Center for Disease Control and Prevention (China CDC). This system selected county and district across China as surveillance points. Currently, the DSPS has included 605 districts and counties comprising 21.1% of all counties and districts in China, and covers a population of 323.8 million (24.3% of the total population). To ensure the representativeness of surveillance points, all selected counties and districts are evenly distributed across different geographical areas with different characteristics.[1] At each surveillance point, all deaths certified by clinical doctors or local CDC professionals are reported to the DSPS in real time using an internet-based reporting system. The data from the DSPS has been widely applied in policy making and disease burden assessment.[2]

Although the DSPS covers a wide range of China territory, and can well represent the Chinese population, many surveillance points especially in western China have small population sizes, and the quality of mortality data were relatively poor which could be indicated by recorded low mortality rate (<5‰). The time-series analyses depend on both good quality of mortality data and adequate daily number of deaths. The small number of daily deaths caused by poor quality of mortality data or small population size may lead to biased associations of temperatures with YLL rates. In the present study, we therefore added surveillance points with good quality of data, which were selected from several provincial mortality surveillance system (Yunnan, Guangdong, Hunan, Zhejiang, and Jilin provinces). To ensure adequate statistical power, surveillance points from both provincial surveillance system and the DSPS were selected if they met anyone or both criteria: a) a population size >200,000, and b) an annual mortality rate >4‰. Under these two criteria, the daily average death count was larger than three, which was consistent with a previous similar study conducted in China.[3]

1.2 Interpolation of daily meteorological factors

We employed the Australian National University Splines (ANUSPLIN) thin plate smoothing software to interpolate the daily nationwide mean temperatures in China. The model was as follows:

$$Temp_i = f\,(lat_i, lon_i) + b \times alt_i + e_i \qquad \text{(Equation 1)}$$

Where $Temp_i$ denotes the daily mean temperature at station i. $f()$ denotes the thin plate function. lat_i, lon_i and alt_i are the latitude, longitude, and altitude for station i. b denotes the coefficient for alt_i, and e_i denotes the error term at station i. Using this method, daily temperatures across China were obtained at a resolution of 0.01°×0.01°. The results of 10-fold cross-validation show good prediction accuracy of the interpolation method for daily mean temperature [$R^2 = 0.96$, root mean squared prediction error (RMSE) = 2.37°C]. In the process of 10-fold cross-validation on the model, we selected only the 15th day in the following months: January 2006, February 2007, March 2008, April 2009, May 2010, July 2011, June 2012, August 2013, September 2014, October 2015, November 2016, and December 2017; this was done as the inclusion of all study days during 2006-2017 would be time-consuming. We employed this method also to interpolate the daily nationwide relative humidity (RH), daily maximum temperatures (TMax) and daily minimum temperatures (TMin) in China. The results of fitting performance show that the R^2 of RH, TMax and TMin were 0.81, 0.94 and 0.94, respectively (Figure S2).

1.3 Daily PM_{10} estimation using a land-use-regression (LUR) model

We selected the PM_{10} as an agency of air quality during the entire study period (2006-2017), and adjusted for it in the DLNM assessing the associations of temperatures with YLL rates. PM_{10} is the only ambient air pollutant obtained during 2006-2012 in this study. Daily average PM_{10} during the 2013-2017 were obtained from the National Urban Air Quality Real-time Publishing Platform (http://106.37.208.233:20035/), which is administrated by the China National Environmental Monitoring Centre. The platform was put in operation since January 2013, and displays real-time concentrations of criteria air pollutants in all state-controlled monitoring stations. Daily average PM_{10} during 2006-2012 were obtained from the China National Environmental Monitoring Centre. The 24-h mean concentrations for PM_{10} were simply averaged from all valid monitoring sties in a city. Since some selected locations were not covered by the air quality monitoring system, we employed a random forest model to assess the daily PM_{10} at each location.

First, we extracted the latitude, longitude and altitude of each included air quality monitoring station across China, and extracted the daily mean temperature, and relative humidity at each monitoring station during the study period from the interpolated data in the appendix section 1.2. We also extracted the population density, length of road, types of land use (farm land, cropland, forest land, water area, and living land) and GDP per capital at each monitoring station using a radius of 1,300 meters, which was chosen based our previous studies.[4 5] The population density data in 2015 were obtained from GeoData Institute in University of Southampton (www.worldpop.org.uk), and the geographic information system (GIS) covariates (geographic map, road density, land use data and GDP per capita) were obtained from the Data Center for Resources and Environmental Sciences (http://www.resdc.cn). Then we established a LUR model implemented by a random forest model to input the above prepared predictors of all air quality monitoring stations and the daily PM_{10} data. One smooth temporal basis function was included in the model to fit the long-term and seasonal trend of PM_{10} concentrations (Figure S3). The results of fitness showed that the R^2 was 73.90%, and the RMSE (Root mean square error) was 16.49 $\mu g/m^3$ (Figure S4). We then extracted the daily mean temperature, relative humidity, latitude, longitude, altitude, population density, road density, land use data and GDP per capita at the center of each selected location, and put them into the established LUR model. Finally, we can obtain the daily mean PM_{10} concentrations during the study period at each location.

1.4 Estimation of the associations of temperatures and TVs with YLL rates

A two-stage approach was employed in our study. In the first stage, we employed a distributed lag nonlinear model (DLNM) to estimate the nonlinear and lag effects of mean temperature on daily YLL rates. The location specific association of temperate with daily YLL rates is entirely defined by a set of parameters namely regression coefficients in the function chosen for representing the associations. The model was as follows:

$$E(YLL_t) = \alpha + \beta_1 T_{1,t,l}(\mathrm{TM}) + \beta_2 T_{2,t,l}(\mathrm{TV}) + ns(\mathrm{RH}, \mathrm{df}) + \beta_3 \mathrm{PM}_{10} + ns(\mathrm{time}_t, \mathrm{df}) + \eta \mathrm{DOW} + e \qquad \text{(Equation 2)}$$

Where t denotes the day of observation, YLL_t denotes the expected YLL rate on day t, which was calculated by dividing the daily observed YLL_t by the total population size (/100,000) in each included location. Therefore, the YLL rate included in the DLNM was the daily YLL in every 100,000 population (YLL/100,000); α denotes the intercept indicative of baseline risk; e denotes a Gaussian error. TM denotes the daily mean temperature. $T_{1,t,l}$ is a matrix obtained by applying the DLNM to mean temperature, and $T_{2,t,l}$ is a matrix obtained by applying the DLNM to TV that is adjusted for as a cofounding factor; β_1 and β_2 denotes the vector of coefficients for $T_{1,t,l}$ and $T_{2,t,l}$, respectively, and l denotes the number of lag days; β_3 denotes the coefficient of daily PM_{10} concentrations; e denotes the error term. To ensure that the meta-analysis can provide meaningful and interpretable results in the second stage, we employed the same function by placing the internal and boundary knots at the same temperatures in all location-specific models. In particular, a quadratic B-spline (*bs*) and natural cubic B-spline (*ns*) were employed to estimate the nonlinear and lagged effects of mean temperature and TV, respectively. In the *bs* function, three internal knots placed at the 10th, 50th and 90th centiles and boundary knots placed at the average minimum and maximum temperatures of location specific temperature distributions were used to model the association of nonlinear curves of mean temperature and TV with the YLL rate. In the *ns* function, an intercept and three internal knots placed at equally spaced values was used to model the lagged effects of temperature and TV. We used a maximum lag period of 21 days to capture the long-term delay of the impact of cold, and also excluded the impact of the harvesting effect. The nonlinear associations of mean temperature and TV with the YLL rate can be interpreted as the effect of the exposure versus a reference which is usually centered on a specific value.[6]

In all epidemiological studies, a basic issue is to control properly for potential confounding. The DLNM model can optimally adjust for confounders including those which change slowly over time (e.g. age, socioeconomic status) expressed as seasonality or long-time trends, day of week (DOW), other meteorological elements, and air pollutants.[7] Additionally, the time-series study design can control for the individual level risk factors such as smoking and alcohol consumption, because these factors at the population level do not vary from day to day, and hence will not influence the short-term effects of temperatures on mortality.[8] Consistent with previous studies,[9 10] we employed a *ns* of time with seven degrees of freedom (*dfs*) per year to control for the seasonal and long-term trends in mortality, a categorical variable to control for the day of the week (DOW), and a linear model to control for the same day PM_{10} concentration. DOW was a dummy variable representing the day of the week, and η was a vector of coefficients. It has been demonstrated that RH is an important contributor to heat stress.[11] We also employed a *ns* to adjust for the potential confounding effect of the present day RH with three *df*s in the *ns* function, which could estimate the nonlinear effects of RH on the YLL rate in the same day.[3] Three internal knots placed at equally spaced values was used to model the nonlinear effects of RH. The family function for DLNM had a Gaussian distribution.

In the second stage, we employed a multivariate meta-analysis method to combine the location-specific 0-21 days' lag in cumulative associations of temperature with the YLL rate.[12]

Multivariate meta-analysis is a method originally developed to pool multiple correlated outcomes in randomized controlled trials. Here it is used to combine the location-specific nonlinear exposure-response curves which are described with function defined by multiple parameters. The multiple parameters obtained from the first stage were used as outcomes for the multivariate meta-analysis, which aims to define an average exposure-response association across the locations, test and quantify the amount of heterogeneity, and further identify the sources of the heterogeneity. Here, in contrast to the original setting of randomized controlled trials, it is not necessary that the parameters are individually interpretable, and the associations is instead characterized through their joint distribution. The multivariate meta-analysis model can be written as follows:

$$\hat{\theta}_i \sim N_k(\theta, S_i + \psi) \qquad \text{(Equation 3)}$$

with $S_i + \psi = \Sigma_i$. The marginal model has independent within-location and between-location components. In the within-location component, the estimated $\hat{\theta}_i$ is assumed to be sampled with error from $N_k(\theta_i, S_i)$, a multivariate normal distribution of dimension k (k is 5 in the present study), where θ_i is the vector of true unknown outcome parameter for location *i*. In the between-location component, $\hat{\theta}_i$ is assumed to be sampled from $N_k(\theta, S_i)$, where ψ is the unknown between-location (co)variance matrix. θ here can be interpreted as the population-average outcome parameters, namely the coefficients of the function s defining the average exposure-response association. A restricted maximum likelihood (REML) method was used to combine the location specific exposure response associations.

The MYT (minimum YLL temperature) was re-centered based on the location-specific MYT identified in the first stage. We specified the combined exposure-response curves with three internal knots (the 10th, 50th and 90th centiles) and two boundary knots (the average minimum and maximum temperatures of location specific temperature distributions). This choice could generate the combined exposure-response curves with the uniform distribution of mean temperatures at national or regional levels. In order to reduce the heterogeneity between locations and uncertainty in the first-stage model, we used a best linear unbiased prediction (BLUP) model to generate the adjusted location-specific exposure-response curves.[12] The BLUP approach made use of a trade-off between the city specific association and the second stage pooled estimation, which could thus provide more precise estimations, especially in cities with small numbers of deaths. We also provided the lag patterns in YLL rates associated with the average extreme cold temperature (2.5th centile), and extreme hot temperature (97.5th centile).

1.5 Calculations of the YLL per death due to temperatures

The YLL per death (YLL*per*) caused by temperatures was calculated using the following two equations:[13]

$$YLL_{tem} = \sum_{i=TMin}^{I=TMax} YLL' \times Freq \qquad \text{(Equation 4)}$$

Where, YLL_{tem} denotes the total YLLs attributable to temperatures in every 100,000 population in a study location. *TMin* denotes the minimum daily temperature, and *TMax* denotes the maximum daily temperature in an included study location. *YLL'* denotes the YLL rate attributable to each temperature, which could be estimated by Equation 2. *Freq* denotes the frequency of each TM in an included study location.

$$YLL_{per} = YLL_{tem}/N \qquad \text{(Equation 5)}$$

Where YLL_{tem} is calculated in Equation 4, and *N* denotes the number of deaths in every 100,000 population in a study location.

1.6 Calculations of the annual mean YLL due to temperatures in China

The annual mean YLL (YLL_{city}) caused by temperatures in each city was calculated using the following equation.

$$YLL_{city} = YLL_{tem} \times Pop/5 \qquad \text{(Equation 6)}$$

Where YLL_{tem} denotes the total YLLs attributable to temperatures in every 100,000 population in a study location, which is calculated in Equation 4. Pop is the annual mean population size (/100000). The number of 5 is the number of years studies (2013-2017). Then we summed the annual mean YLLs attributable to temperatures in all cities, and obtained the annual mean YLLs across China. All codes were available in https://github.com/gztt2002/YLL-of-Tm-and-Tv.

2.Supplementary tables

Table S1. General characteristics of included 364 locations in China

	Mean (SD)	Minimum	25th centile	Median	75th centile	Maximum
Population size (×10,000)	56.9 (31.9)	2.3	33.7	49.2	74.1	207.2
Percentage of male (%)	51.1 (1.2)	48.4	50.4	51.0	51.7	56.8
Percentage of elderly (%)	10.4 (2.3)	3.5	8.8	10.4	11.7	18.7
GDP per capital (×1000 yuan)	3.3 (2.5)	0.4	1.6	2.7	4.2	21.5
Urbanization rate (%)	52.8 (24.4)	6.4	34.5	46.9	67.7	100.0
Average years of education	10.0 (1.5)	5.2	9.1	9.8	10.7	16.7

Table S2. YLL per death due to different components of temperature across 364 locations in China

	YLL per death (years, 95%CI)			
	Extreme cold	Moderate cold	Moderate heat	Extreme heat
Causes of death				
Total mortality	0.14 (0.12, 0.16)	0.84 (0.52, 1.16)	0.01 (0.01, 0.02)	0.03 (0.02, 0.03)
Cardiovascular disease	0.17 (0.15, 0.19)	0.93 (0.59, 1.27)	0.01 (0.01, 0.02)	0.03 (0.02, 0.04)
Cerebrovascular disease	0.17 (0.14, 0.20)	1.16 (0.70, 1.62)	0.01 (0.00, 0.02)	0.03 (0.02, 0.04)
Respiratory disease	0.14 (0.11, 0.18)	0.28 (0.13, 0.44)	0.01 (0.01, 0.02)	0.02 (0.01, 0.03)
Geographic regions				
Northern	0.05 (0.01, 0.10)	0.51 (-0.84, 1.86)	0.03 (-0.12, 0.19)	0.04 (-0.03, 0.11)
Central	0.15 (0.12, 0.17)	1.14 (0.64, 1.64)	0.02 (0.01, 0.03)	0.04 (0.03, 0.05)
Southern	0.12 (0.09, 0.14)	1.05 (0.59, 1.50)	0.01 (0.00, 0.01)	0.02 (0.01, 0.03)
Age of death (Years)				
0-64	0.2 (0.15, 0.25)	0.84 (0.29, 1.39)	0.01 (0.00, 0.01)	0.03 (0.02, 0.05)
65-74	0.16 (0.13, 0.2)	1.11 (0.60, 1.62)	0.01 (0.00, 0.02)	0.02 (0.01, 0.03)
≥75	0.12 (0.11, 0.14)	0.69 (0.51, 0.87)	0.01 (0.01, 0.02)	0.02 (0.02, 0.03)
Sex				
Male	0.15 (0.12, 0.17)	0.98 (0.57, 1.38)	0.01 (0.00, 0.01)	0.02 (0.02, 0.03)
Female	0.14 (0.12, 0.17)	0.68 (0.28, 1.07)	0.02 (0.01, 0.03)	0.03 (0.02, 0.04)

3.Supplementary figures

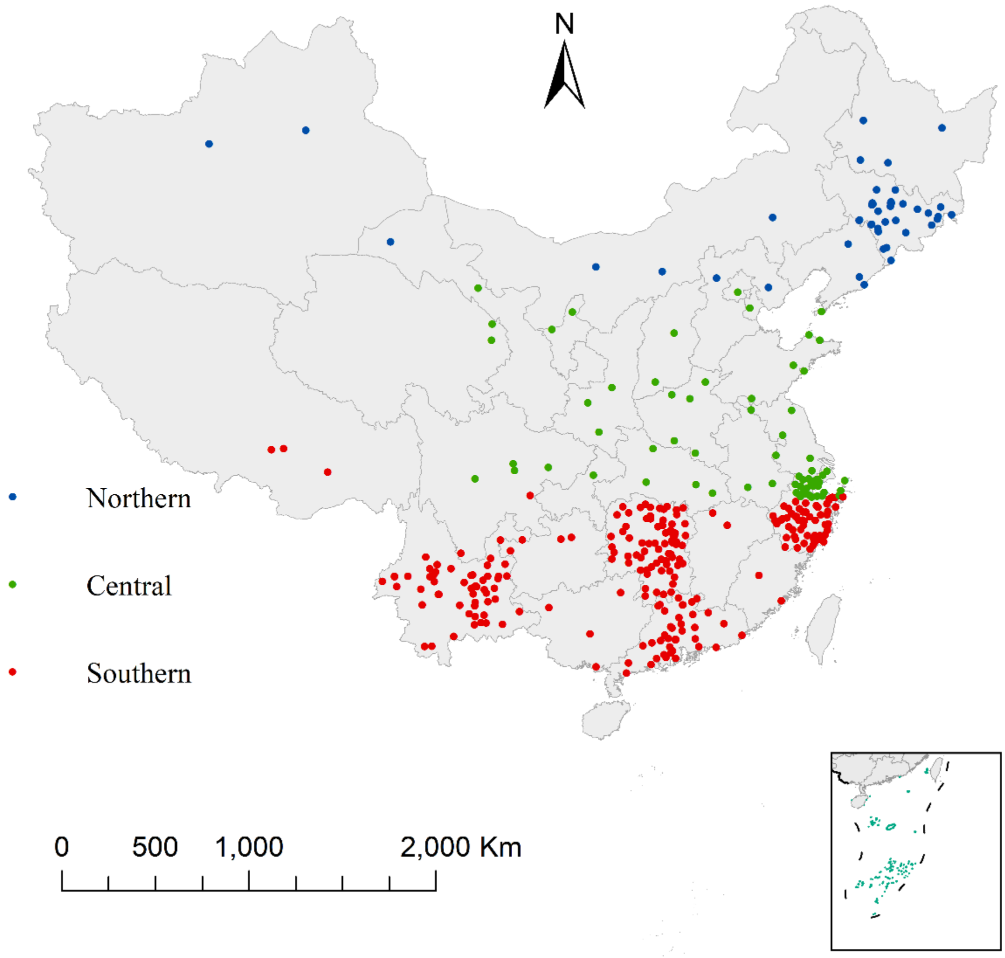

Figure S1. The geographic distribution of 364 locations in China

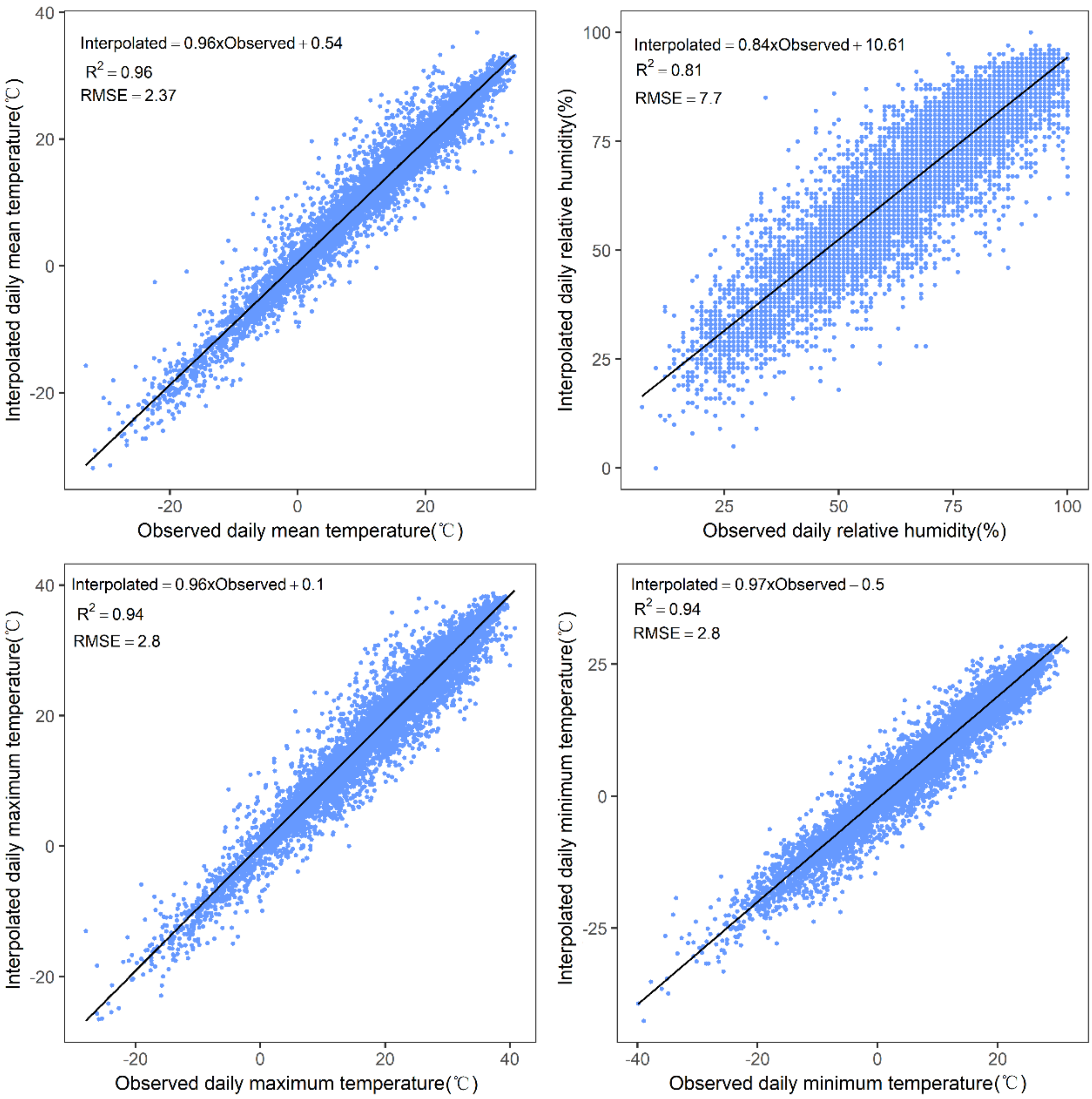

Figure S2. Scatter plot of 10-fold cross-validation of interpolated meteorological factors in 698 weather stations in China

Note: Since 10-fold cross-validation of all study days during 2006-2017 was time consuming, we only selected the 15th day in the following months to test the fitting performance of the model: January 2006, February 2007, March 2008, April 2009, May 2010, July 2011, June 2012, August 2013, September 2014, October 2015, November 2016, and December 2017.

RMSE: root mean squared prediction error (°C)

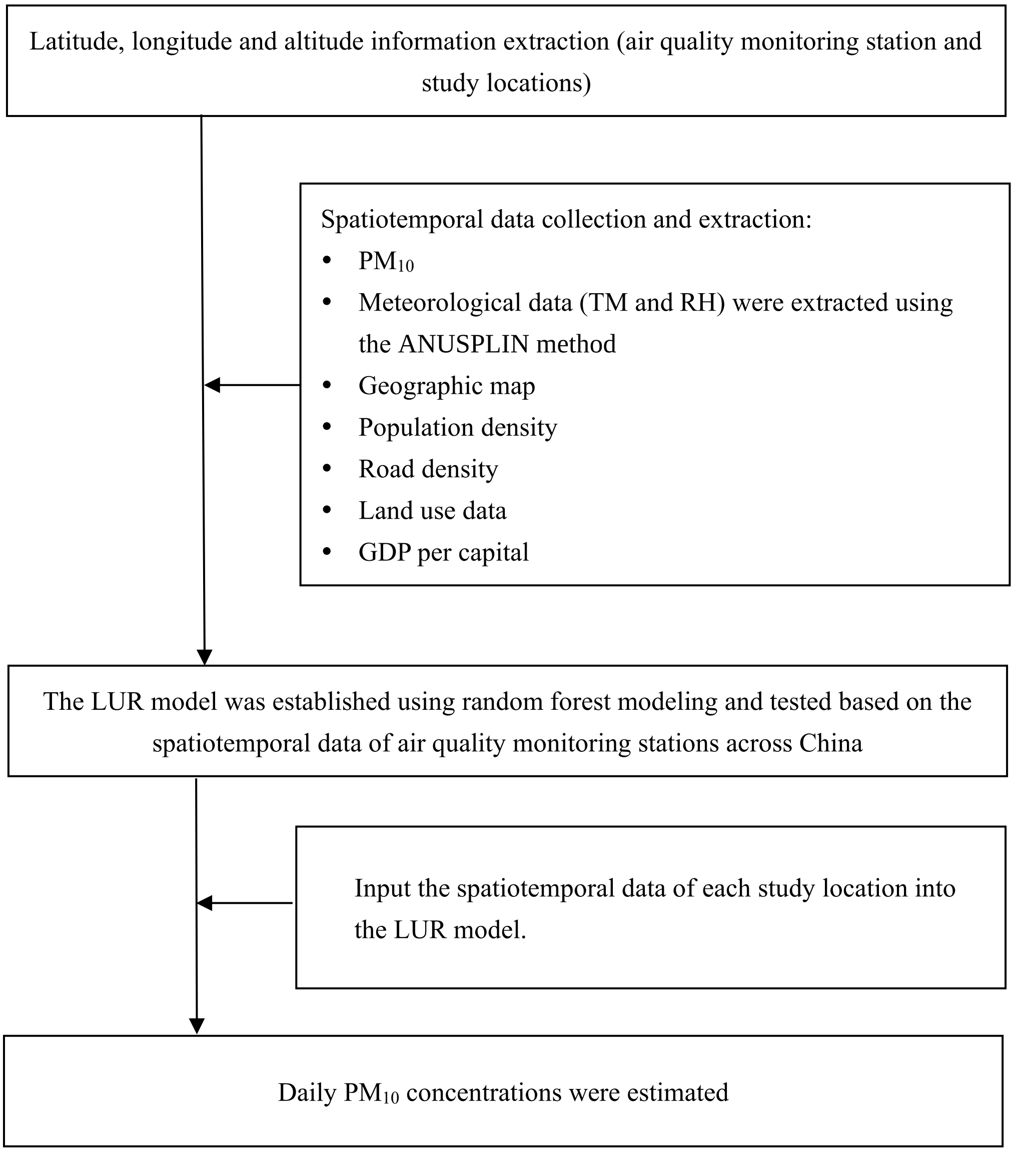

Figure S3. The process of daily PM_{10} concentration estimation

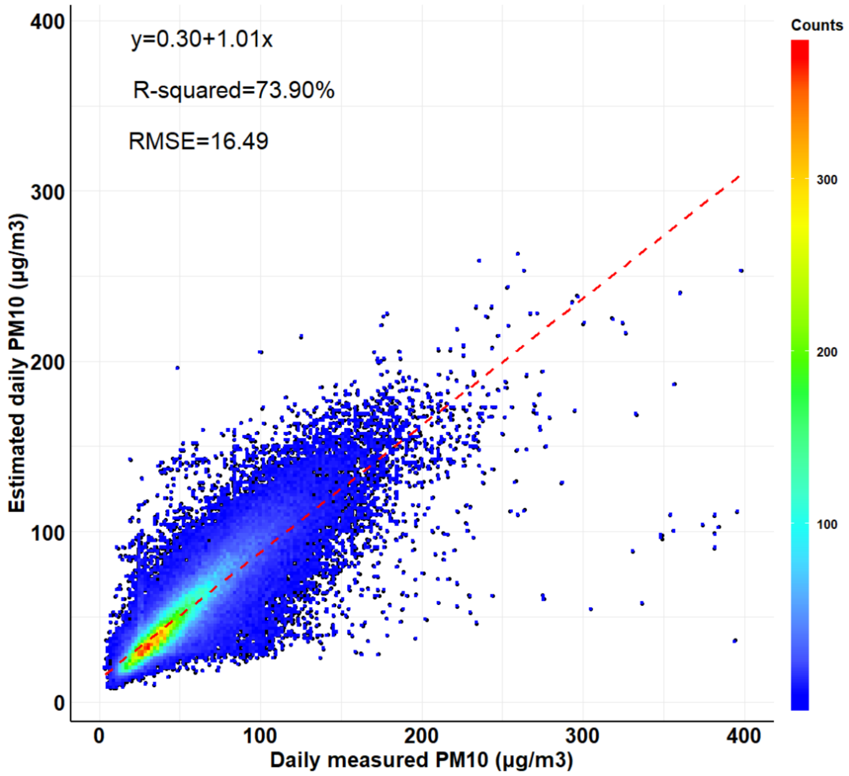

Figure S4. Performance and validation of estimating the daily PM_{10} concentration in China

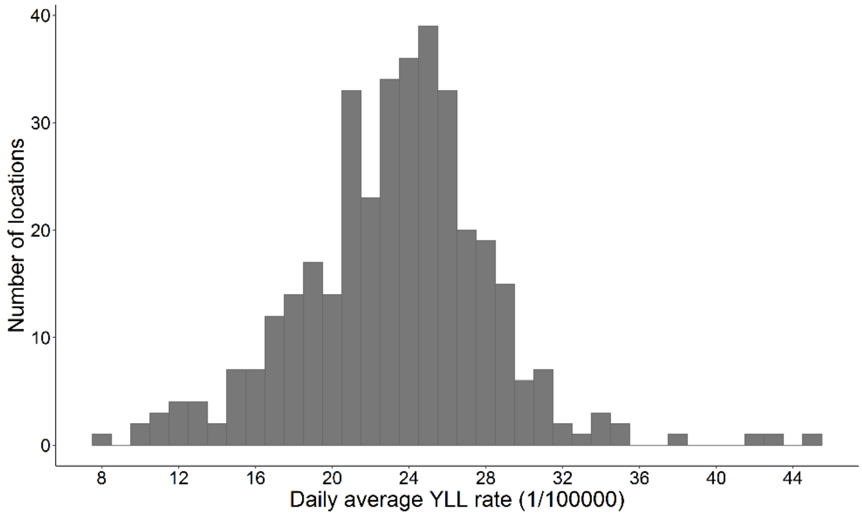

Figure S5. The distribution of daily average YLL rate (/100,000 population) in 364 locations in China

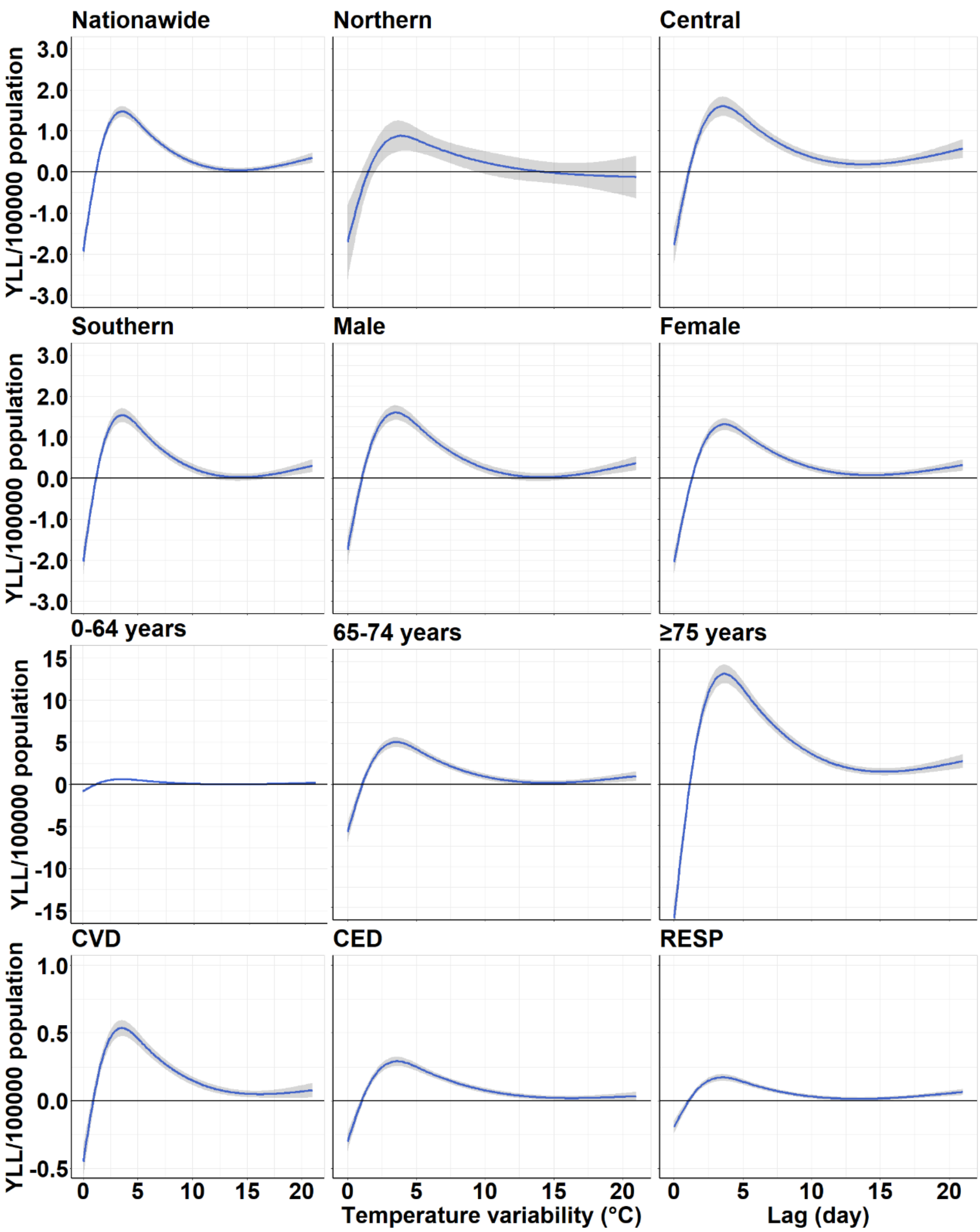

Figure S6. Overall lag structure in effects of extreme cold temperature on daily YLL rates in all 364 locations and in different subgroups in China

Effects were defined as the risks at the mean of the 2.5th centile of temperature distributions compared with the estimated MYT

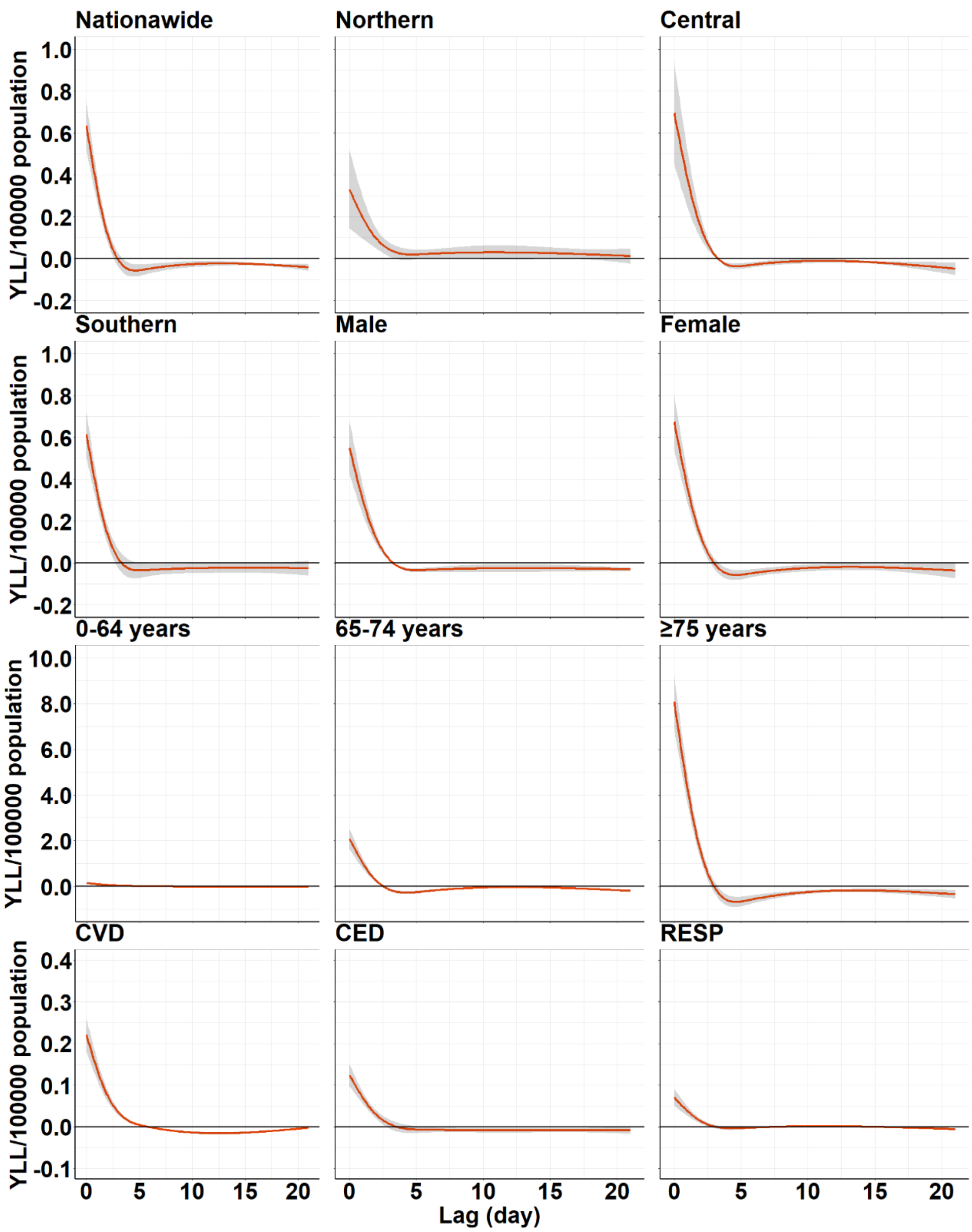

Figure S7. Overall lag structure in effects of extreme hot temperature on daily YLL rates in all 364 locations and in different subgroups in China

Effects were defined as the risks at the mean of the 97.5th centile of temperature distributions compared with the estimated MYT

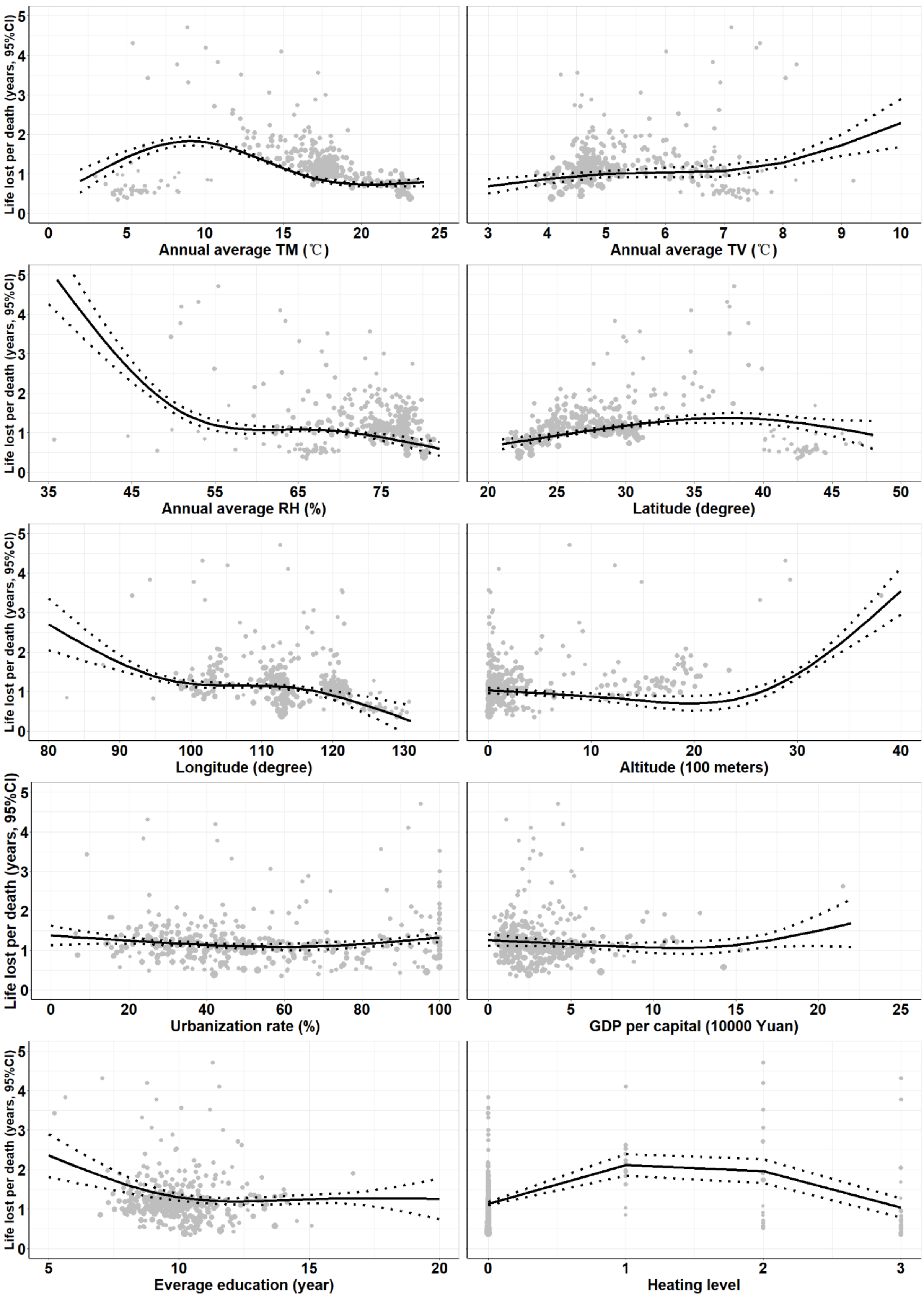

Figure S8. Univariable meta-regression results of the modification effects of city level characteristics on the YLL per death caused by temperatures in 364 locations in China

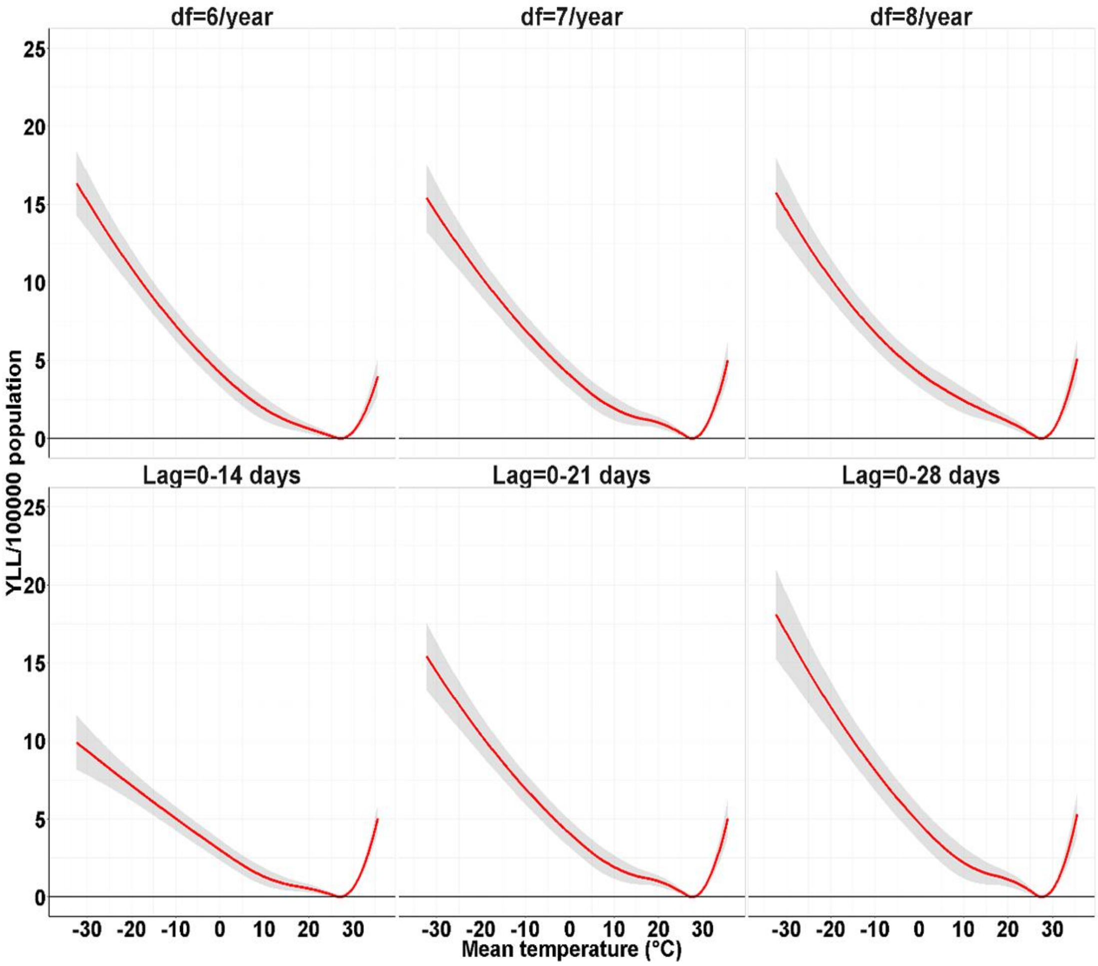

Figure S9. Sensitivity analyses on the impacts of lag days and df (/year) on the nationwide exposure-response relationship between temperature and YLL rate in China

References

1 Liu, S., Wu, X., Lopez, A.D., et al. (2016). An integrated national mortality surveillance system for death registration and mortality surveillance, China. Bull World Health Organ. 94, 46-57.

2 Zhou, M., Wang, H., Zhu, J., et al. (2016). Cause-specific mortality for 240 causes in China during 1990–2013: a systematic subnational analysis for the Global Burden of Disease Study 2013. Lancet. 387, 251-272.

3 Chen, R., Yin, P., Wang, L., et al. (2018). Association between ambient temperature and mortality risk and burden: time series study in 272 main Chinese cities. BMJ. 363, k4306.

4 Liu T, Chen X, Xu Y, et al. (2019) Gut microbiota partially mediates the effects of fine particulate matter on type 2 diabetes: Evidence from a population-based epidemiological study. Environ Int 130: 104882.

5 Liu T, Xiao J, Zeng W, et al. (2019) A spatiotemporal land-use-regression model to assess individual level long-term exposure to ambient fine particulate matters. MethodsX 6: 2101-05.

6 Cao J, Valois M-F, Goldberg MS. (2006) An S-Plus function to calculate relative risks and adjusted means for regression models using natural splines. Comput Methods Programs Biomed 84(1): 58-62.

7 Armstrong B. (2006) Models for the relationship between ambient temperature and daily mortality. Epidemiology 2006; 17(6): 624-31.

8 Schwartz J, Spix C, Touloumi G, et al. (1996) Methodological issues in studies of air pollution and daily counts of deaths or hospital admissions. J Epidemiol Community Health 50 Suppl 1: S3-11.

9 Gasparrini, A., Guo, Y., Hashizume, M., et al. (2015). Mortality risk attributable to high and low ambient temperature: A multicountry observational study. Lancet. 386, 369-375.

10 Ma, W., Wang, L., Lin, H., et al. (2015). The temperature–mortality relationship in China: An analysis

from 66 Chinese communities. Environ Res. 137, 72-7.

11 Armstrong B, Sera F, Vicedo-Cabrera AM, et al. (2019) The role of humidity in associations of high temperature with mortality: a multicountry, multicity study. Environ Health Perspect 2019; 127(9): 97007.

12 Gasparrini, A., Armstrong, B., Kenward, M.G.. (2012). Multivariate meta-analysis for non-linear and other multi-parameter associations. Stat Med. 31, 3821-39.

13 Majdan M, Plancikova D, Maas A, et al. (2017) Years of life lost due to traumatic brain injury in Europe: A cross-sectional analysis of 16 countries. Plos Med 2017;14(7): e1002331.