Association between temperature variation and cardiovascular disease hospitalization: A case study in Shanxi, China

Miao Cai

Department of Epidemiology and Biostatistics,
College for Public Health and Social Justice,
Saint Louis University, Saint Louis, MO 63103
miao.cai@slu.edu

A Draft on May 15, 2019

Abstract

Objective: Previous studies primarily focused on the association between extreme temperature and health outcomes, but few explored the association between temperature change and health outcomes. This paper aims to explore the association between temperature variation and cardiovascular disease (CVD) patient hospitalization based on a province-wide sample from Shanxi, China.

Methods: A total of 1.44 million CVD hospital admissions were identified from 175 secondary and tertiary hospitals in Shanxi from 2015 to 2017. We decomposed temperature variation into horizontal variation (HV) and vertical variation (VV). A traditional Poisson model, a hierarchical Poisson model, generalized additive Poisson models, and a spatial lag Poisson model were used to examine the association between the two sources of temperature variation and the number of CVD hospitalization after adjusting for the average temperature, GDP per capita, the percent of female and rural population, and the total population.

Results: Univariate Moran's I revealed significant spatial clustering effects of the outcome and all predictor variables. The value and variance of HV were consistently lower than those of VV in all 117 administrative districts. The incidence rate ratio estimate for HV was reduced from 1.072 (95% CI: [1.06, 1.083]) in the mixed-effects Poisson model to 1.023 (95% CI: [1.011, 1.035]) in the spatial lag Poisson model. It decreased from 1.015 (95% CI: [1.008, 1.023]) in the mixed-effects Poisson model to 1.003 (95% CI: [0.995, 1.01]) in the spatial lag Poisson model for VV.

Conclusion: Temperature variation across multiple days is a robust risk factor for CVD hospitalization.

Contents

List of Tables

List of Figures

1 Introduction

The emission of anthropogenic greenhouse gases has increased the frequency, intensity, and duration of extreme weather events globally (Ye et al. 2011; Perkins, Alexander, and Nairn 2012). It was estimated that 12.6 million deaths were attributable to environmental factors that can be modified. These environmental factors were either driven by or associated with climate change (Prüss-Üstün and Neira 2016; Watts et al. 2017). These climate events have been reported to have posed an emerging threat to cardiovascular disease (CVD) patients, including those diagnosed with hypertension, myocardial infarction, heart attack, and others (Basu and Samet 2002; Braga, Zanobetti, and Schwartz 2002). CVD was the leading cause of death worldwide in 2016, killing over 15.2 million people, which accounted for 26.7% of all deaths (The World Health Organization 2018). Specifically, CVD was the primary cause of death for both urban and rural population in China in 2016, leading to 2,903.1 and 2,728.7 deaths per one million people (National Health and Family Planning Commission, The People's Republic of China 2018; Jiang et al. 2015).

Previous studies suggested that long-time exposure in extreme temperature, defined as temperature consistently higher or lower than a threshold, were associated with higher morbidity and mortality (Ryti, Guo, and Jaakkola 2015). However, these studies ignored the temperature variations across multiple periods. In contrast with long-time exposure to extreme temperature, temperature variation, such as the range between the maximum and minimum temperature within a day, is potentially associated with CVD occurrence and mortality by increasing patients' heart rates, oxygen uptake, and cardiovascular workload (Cheng et al. 2014; Qiu et al. 2013; Shinkawa et al. 1990). In contrast, this topic has been much less studied compared with extreme temperature. Besides, previous studies used either traditional statistical models or hierarchical models to explore the association between extreme weather and CVD outcomes. These models ignored the spatial clustering and autocorrelation in real life data and may lead to biased estimates.

Considering the high prevalence of CVD coupled with an unclear understanding of the association between temperature variation and CVD outcomes, it is essential to have a more comprehensive understanding on the effect of temperature variation on CVD patient's health

outcomes, which contribute to reducing CVD related disease of burden (Li et al. 2015). In this study, we propose to examine this association using two measures of temperature variation (horizontal and vertical variation) based on inpatient data from Shanxi, China. Shanxi is a North China province that has a wide range in latitude and complex and different types of landforms, which results in very different temperature patterns between the north and the south, between winter and summer, and between day and night. This provides us a perfect opportunity to study the association between temperature variation and CVD outcomes across 117 administrative districts (ADs) in Shanxi. This study promotes the understanding of the association between weather changes and CVD hospitalization and can help reduce CVD hospitalizations attributable to weather changes.

2 Motivation for research

2.1 CVD is the leading cause of death worldwide and in China.

Recent years have witnessed a shift of disease spectrum to non-communicable disease (NCD). NCD now accounts for over a half of global mortality and morbidity, with a significant increase in the proportion compared with that 30 years ago.12 CVD consists of over 50% of all NCD deaths, and almost 70% of these deaths occur in low- and middle-income countries (Benziger, Roth, and Moran 2016). According to the latest healthcare report by the Chinese government, CVD was the primary cause of mortality in both urban and rural areas of China in 2016.6 The yearly number of CVD deaths increased from 25% in 1990 to 40% in 2010, and this trend is predicted to continue until at least 2030, leaving CVD as a pressing public health concern in China (Jiang et al. 2015). Therefore, exploring public health factors associated with the incidence and disease burden of CVD patients can benefit the giant population in China.

2.2 Dramatic weather changes and human activities

People's activities, such as vehicle emission and burning of fossil fuels, contribute to an increase of global surface temperature. It is estimated that the average global surface temperature increased by 0.26 Celsius degree per decade from 1956 to 2005, compared to only

0.07 Celsius degree per decade between 1906 and 2005 (Ye et al. 2011). In contrast with hot and cold weather effects, temperature variation increases heart workload by increasing blood pressure, heart rate and oxygen intake (Cheng et al. 2014). These physiological changes led by temperature variation can cause acute myocardial infarction, cardiac failure, and stroke, which pose a threat to people's risk of death and quality of life (Lim, Hong, and Kim 2012). These potential consequences of drastic temperature change are especially significant among older and CVD patients since they are less capable of adapting to external temperature change and maintain internal homeostasis (Lim et al. 2013).

2.3 A potential association between temperature variation and CVD outcomes

Although plenty of studies have been published on the association between climate change and CVD outcomes, very few papers focused on temperature variation or change. In contrast with long-time exposure to extreme temperature, temperature variation, such as the range between the maximum and minimum temperature within a day, has been reported to be associated with CVD occurrence and mortality by increasing patients' heart rates, oxygen uptake and cardiovascular workload (Cheng et al. 2014, 2014; Qiu et al. 2013; Shinkawa et al. 1990; Ryti, Guo, and Jaakkola 2015). Besides, it is suggested that drastic temperature variation could impact health adversely through biological pathways, such as deteriorating people's cardiovascular, nervous and immune system (Liang et al. 2008). However, the influence of temperature variation is much less investigated and understood compared with the abundant literature on extreme temperature and health impacts.

2.4 Less study accounts for spatial correlation

Previous studies on weather and CVD outcomes either used traditional statistical models that assume the observations were independent (L. Yang et al. 2015; J. Yang et al. 2015), or hierarchical models that assume the observations were conditionally independent (Sartini et al. 2017; Phung et al. 2016). These studies created value in understanding the association between temperature variation and CVD outcomes. However, these studies failed to account for the fact that temperature data that were generated in geographically nearby areas were

similar. Ignoring this spatial clustering and auto-correlation nature of temperature data will throw away a piece of crucial information, and can therefore lead to biased estimates that reduce the validity of scientific findings (Blangiardo and Cameletti 2015).

3 Literature review

3.1 Extreme temperature and health outcomes

Numerous studies, including both empirical studies and systematic reviews, have been published on the association between exposure in extreme temperature and the health outcomes of CVD patients. For example, Ye et al. (2011) reviewed epidemiological evidence on the association between both heat waves and cold spells on noncommunicable diseases, and they determined a significant short-term effect of ambient temperature on all-cause and cause-specific morbidities. Basu et al. (2012) conducted meta-analyses to combine conditional logistic regressions in different climate zones. They found that high ambient temperature had the same-day effects on emergency room admissions for several outcomes. Gronlund et al. (2014) found that heat waves were associated with 3% (95% CI: [2%, 4%]) increased elderly hospitalization in the subsequent eight days in the United States, especially the hospitalization attributable to renal causes (15%, 95% CI: [9%, 21%]). A systematic review and meta-analysis by Ryti, Guo, and Jaakkola (2015) summarised that cold spells were associated with increased mortality rates around the world, although the studies they included were substantially heterogeneous.

Specifically for CVD, a number of studies also reported the association between extreme temperature and CVD health outcomes. For example, Pan, Li, and Tsai (1995) found a U-shaped relation between temperature and the mortality of coronary artery disease and cerebral infarction among Chinese in Taiwan from 1981 to 1991, which suggested that both low and high temperature can increase CVD mortality. Lin et al. (2009) reported that each Celsius degree increase above the temperature effect curve (29 - 36°C) was associated with 2.7% to 3.1% increase in respiratory disease hospitalization on the same day in New York City. On the other hand, a few studies also reported a null or insignificant association between extreme temperature. For example, Michelozzi et al. (2009) found that one Celsius

degree increase in the maximum temperature above a threshold was associated with 4.5% (95% CI: [1.9%, 7.3%]) increase in respiratory admissions, but no significant association with CVD admissions. Cheng and Su (2010) reviewed the effect of both extreme hot and cold temperature on CVD outcomes in both epidemiological and clinical perspective. They found that extreme weather events, including both hot and cold waves, have been widely studied and can be explained in clinical pathways. However, they claimed that the effect of extreme climate change on CVD outcomes was hard to predict due to unmeasured confounders.

3.2 Diurnal temperature range and health outcomes

Diurnal temperature range (DTR) is defined as the difference between the maximum and minimum temperature within one day (Cheng et al. 2014). It is an important meteorological variable that measures the stability of weather. A systematic review by Cheng et al. (2014) found that DTR was significantly associated with both mortality and morbidity, particularly among CVD and respiratory disease patients. However, the vulnerable groups, lag time, and threshold of DTR were still inconsistent and needed to be explored further. Onozuka and Hagihara (2017) reported an association between DTR and out-of-hospital cardiac arrest in six major prefectures in Japan from 2005 to 2013 in both cold and hot temperature days. Zheng et al. (2016) found that a short-term increase in DTR was significantly associated with CVD cause-specific emergency room admissions in Beijing, with gender, age, and season as potential modifiers. Another study by Yang et al. (2018) suggested that one Celsius degree increase in DTR was associated with an overall 0.66% ([0.28-1.05%]), 0.12% ([−0.26-0.51%]), and 0.67% ([0.26-1.07%]) increase in stroke mortality on all, hot, and cold days in 16 provincial capital cities in China from 2007 to 2013.

4 Research question and hypothesis

The research question in this study is whether temperature variation is associated with CVD hospitalization based on the sample from Shanxi, China? We hypothesize that more regional temperature variation is associated with a higher rate of CVD hospitalizations after controlling for the patient- and regional-level variables, and spatial autocorrelation.
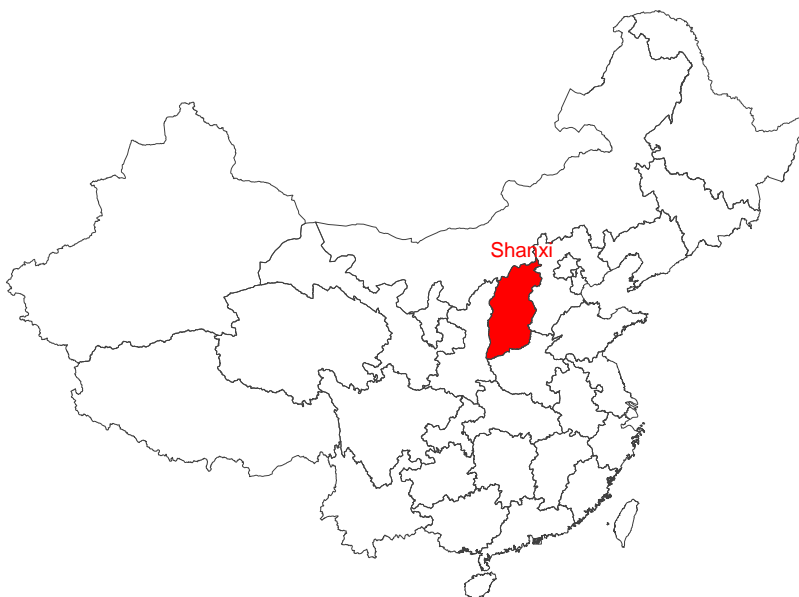
Figure 1: The location of Shanxi Province in China

5 Methods

5.1 Data source

Shanxi is located in North China region (Figure 1), with over 36 million residents living in 117 ADs (Bureau of Statistics, Shanxi 2018), with an average of about 300,000 residents living in one AD. These ADs are either districts or counties. Districts are typically urban cities with high population density and urbanization level, while counties are in rural areas with low population density and urbanization level. As Shanxi has a wide range in latitude and complex and different types of landforms, it has distinct four seasons, very different temperature patterns between north and south Shanxi, between winter and summer, and between day and night (Wikipedia contributors 2019). The temperature data provided by the China Meteorological Administration suggested that the average temperature difference within a day in Shanxi is 12°C, while the average temperature difference between summer and winter is 28.1°C.
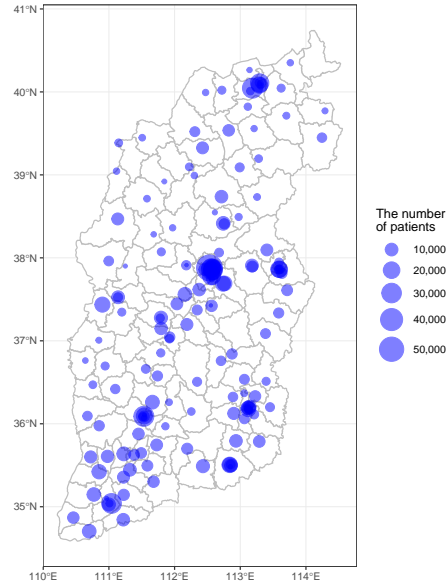
Figure 2: Geographic distribution of 175 hospitals in Shanxi, China

We used three different databases in Shanxi province, China: Shanxi inpatient database provided by the Health and Family Planning Commission in Shanxi, AD level weather data provided by the China Meteorological Administration, and socio-demographic data from Shanxi Statistical Yearbook (Bureau of Statistics, Shanxi 2018). The Shanxi inpatient database includes 175 major secondary and tertiary hospitals that cover most of the CVD hospitalization in Shanxi, China. The geographical distribution of the sample 175 hospitals can be seen in Figure 2.

Using the 10th revision of the International Statistical Classification of Diseases (ICD-10), we identified around 1.44 million CVD inpatients with the primary diagnosis ICD-10 code that begins with "I" from the Shanxi inpatient database from 2015 to 2017. These CVD patients included heart, hypertensive, cerebrovascular, arteries and veins diseases. The ICD-10 coding in Chinese electronic medical records originated from the World Health Organization and may be subjected to adjustment for local diseases. These coding has been validated and can be used for large population-based studies (Kurmi et al. 2016). Since the total number of CVD hospitalization is very small, we aggregated it on a monthly basis.

We also obtained daily maximum, minimum, and average temperature, as well as air quality index in each ADs in Shanxi during the same period. Sociodemographic variables, including the total population, Gross Domestic Product (GDP) per capita, the percent of female, and the percent of the rural population were obtained from the Shanxi Statistical Yearbook from 2015 to 2017 (Bureau of Statistics, Shanxi 2018). These three data sets (CVD inpatient data, weather data, and sociodemographic data) were then merged using a unique AD identifier. After excluding areas with missing data, a total of 109 administrative districts (ADs) were included in this study. Each AD had patients data in each 12 months from 2015 to 2017, which resulted in $109 * 12 * 3 = 3,924$ observations as our final sample.

5.2 Variables

5.2.1 Temperature variation

We decomposed the daily temperature variation in our study into two sources: vertical variation and horizontal variation. As demonstrated in Equation (1), horizontal variation (HV) was defined as the absolute difference between the average temperature in the present day and that in the previous day. On the other hand, vertical variation (VV), also known as the diurnal range in Equation (2), was defined as the absolute difference between the maximum and the minimum temperature during a day (Wang et al. 2013). Compared to horizontal temperature variation that measures temperature changes across multiple days, vertical variation captures the temperature variation within a day. Different lags from 0 to 10 have been used in literature to calculate temperature lagging effect (Cheng et al. 2014). In this study, we calculated the mean of temperature, horizontal variation, and vertical variation on a monthly basis with a lag of seven days, which has been commonly used in previous literature (Cheng et al. 2014; Onozuka and Hagihara 2017; Hu et al. 2019).

$$HV = \sum_{i=d-7}^{d} |T_i - T_{i-1}| \tag{1}$$

$$VV = T_{max} - T_{min} \tag{2}$$

5.2.2 Covariates

Based on previous studies and the availability of data, we added four potential confounding variables, average temperature, the natural logarithm of GDP per capita in Chinese Yuan (1 Yuan equals around 0.15 United States Dollar as of May 15, 2019), the percent of female, and the percent of the rural population, as covariates in this study. Since the GDP per capita has large variance and an extremely right-skewed distribution and one unit change in it would have a very nominal effect on the outcome variable, we took the natural logarithm of it. The total population was used as an auxiliary variable that was accounted for but did not have parameter estimates. This auxiliary variable has also been known as the offset term in Poisson models.

5.3 Statistical models

Since the dependent variable in this study was the number of patients in each AD, it must be non-negative and integers. Besides, exploratory data analysis suggested that it had a highly right-skewed distribution. Therefore, multiple linear regression with a normally distributed error term was unlikely to give rise to this data. Instead, a Poisson process that could generate non-negative, integer, and right-skewed count data was used to model the data in this study. We used five sets of Poisson regressions in this study: Poisson regressions with and without smoothing splines, hierarchical Poisson regressions accounting for the district- or county-level individual effect with and without smoothing splines, and a spatially weighted Poisson regression.

5.3.1 Non-spatial Poisson model

The non-spatial Poisson model in this study was parameterized as Equation (3).

$$
\begin{aligned}
Y_i &\sim \text{Poisson}(\text{Pop}_i * \lambda_i) \\
\log(\lambda_i) &= \beta_0 + \sum_k \beta \mathbf{X}
\end{aligned}
\tag{3}
$$

Where Y_i is the number of CVD inpatients as the outcome variable, Pop_i is the total population as the auxiliary variable, which is also known as an offset term, λ_i is the rate parameter of Poisson distribution, which is both the mean and variance of a Poisson distribution, and $\mathbf{X}$ is the predictor matrix including horizontal temperature variation (HV), vertical temperature variation (VV), monthly average temperature, the natural logarithm of GDP per capita, the percent of female, and the percent of the rural population. In order to test the non-linear effect of temperature variation on CVD hospitalization, we also conducted a generalized additive Poisson regression that used two smoothing splines for two temperature variation besides this non-spatial Poisson model.

5.3.2 Non-spatial Hierarchical Poisson model

Since each AD was measured for 36 times (12 months a year from 2015 to 2017) and each AD can have its own characteristics, the assumption that the observations were independent in Equation (3) was not met in this study. Therefore, we used a hierarchical Poisson regression that assumes conditional independence of observations parameterized in Equation (4).

$$
\begin{aligned}
Y_{ij} &\sim \text{Poisson}(\text{Pop}_{ij} * \lambda_{ij}) \\
\log(\lambda_{ij}) &= \beta_{0j} + \beta_{1j}HV_{ij} + \beta_{2j}VV_{ij} + \sum_{k-2}\beta\mathbf{X} + u_{0j} \\
\beta_{0j} &\sim N(\mu_0, \sigma_0^2) \\
\beta_{01} &\sim N(\mu_1, \sigma_1^2) \\
\beta_{02} &\sim N(\mu_2, \sigma_2^2)
\end{aligned}
\tag{4}
$$

Where β_{0j} is the random intercepts for each AD j, β_{1j} and β_{2j} are the random slopes of HV_j and VV_j for each AD j. Likewise, we also conducted a hierarchical generalized additive Poisson regression that used smoothing splines for the two sources of temperature variation apart from this hierarchical Poisson model.

5.3.3 Spatial lag Poisson model

Although the random effects models proposed in Equation (4) accounted for the clustering nature of different ADs, they failed to recognize the spatial structure and clustering nature of these ADs. Therefore, we proposed a spatial lag model parameterized in Equation (5).

$$
\begin{aligned}
Y_i &\sim \text{Poisson}(\text{Pop}_i * \lambda_i) \\
\log(\lambda_i) &= \beta_0 + \sum_k \beta \mathbf{X} + \rho \mathbf{W} y
\end{aligned}
\tag{5}
$$

Where ρ is an autoregressive parameter, $\mathbf{W}$ is a spatial weighting matrix that measures the relative distance between different ADs. We did not use a spatial error Poisson model or a geographically weight Poisson model due to the underdevelopment of software packages that could be used to implement them (Simoes and Natário 2016; Glaser 2017).

All analyses were performed in the statistical computing environment `R 3.6.0` (R Core Team 2018). Specifically, the spatial objects were handled using the `sf` package (Pebesma 2018), data manipulation and visualization were performed using the `tidyverse` package (Wickham 2017), non-spatial Poisson, hierarchical Poisson, non-hierarchical and hierarchical generalized additive Poisson, and spatially weighted Poisson regression were conducted using `glm` function from base `R` (R Core Team 2018), `lme4` (Bates et al. 2015), `mgcv` (Wood 2017), and `INLA` packages (Bakka et al. 2018; Rue et al. 2017), respectively.

6 Results

6.1 Thematic Maps of the dependent variable

Figure 3 shows the choropleth map of the number of patients in each AD in Shanxi, China. There were two major urban cores in Shanxi province. The first one was in the center of the province concentrated around Taiyuan while the other one was in the south of the province around Yuncheng and Linfen. These three cities were also the top three cities with the most population in Shanxi.
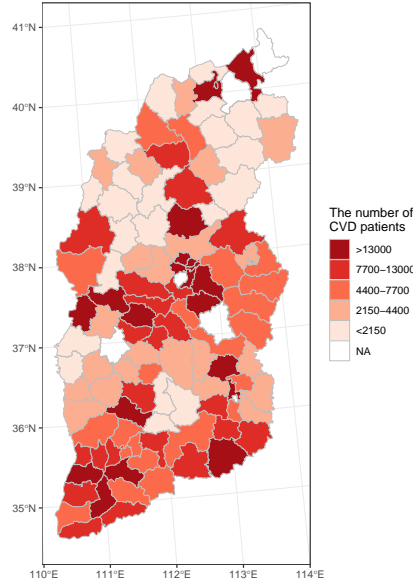
Figure 3: Choropleth map patient distribution in each AD in Shanxi, China

6.2 Spatial Descriptive Statistics

In Figure 4, each AD has two lines, with the red lines representing the mean maximum temperature in a month while blue lines representing the mean minimum temperature in a month. This figure demonstrates significant differences in temperature patterns across the 117 districts and counties, across different time periods and between the maximum and minimum temperature.

Figure 5 demonstrates the mean horizontal (red lines) and vertical variation (blue lines) in the 36 months in 117 ADs in Shanxi China from 2015 to 2017. It shows that HV was consistently lower than those of VV in all 117 ADs. Besides, the variance of HV across different ADs was also consistently smaller than that of VV.

6.3 Spatial Autocorrelation

Table 1 shows univariate Moran's I for the outcome, predictor, and auxiliary variables (total population). The univariate Moran's I results demonstrated that all these variables were significantly spatially autocorrelated.
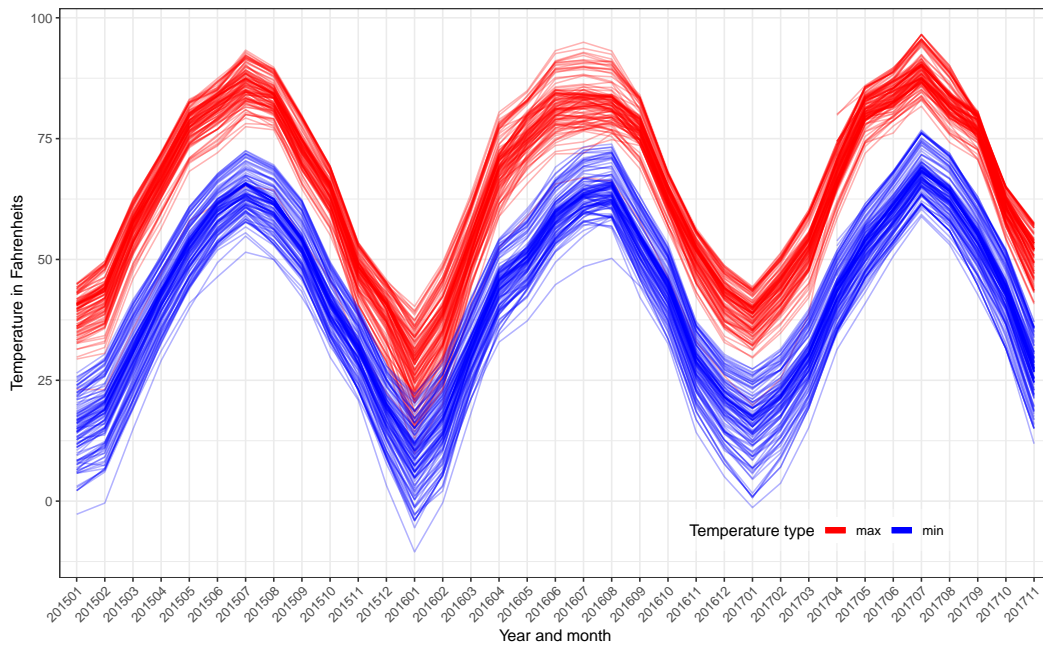
Figure 4: Monthly average maximum and minimum temperature in 117 ADs in Shanxi, 2015 - 2017
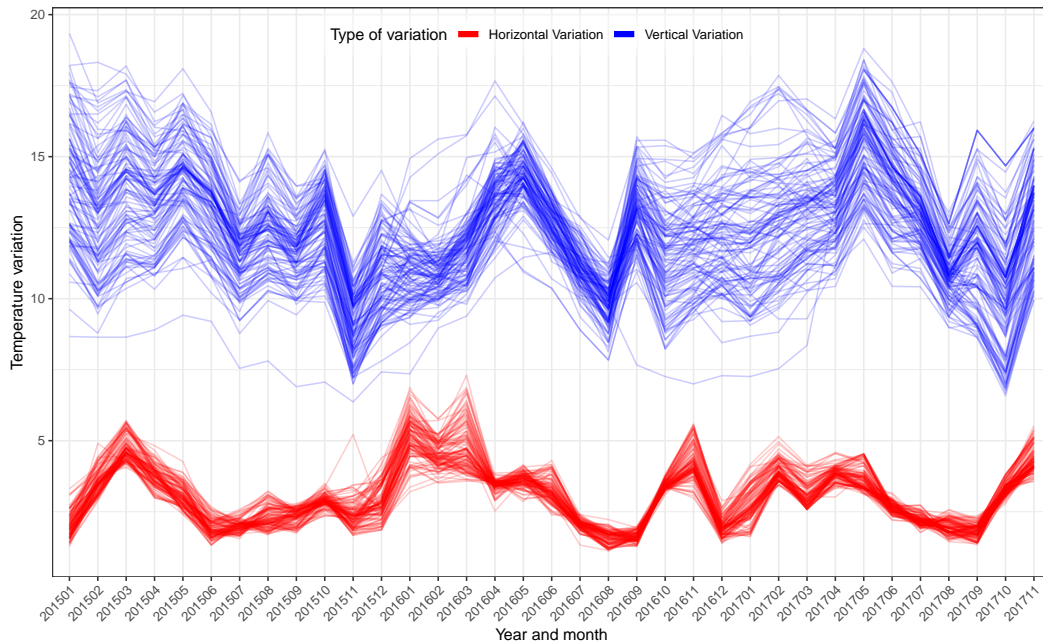
Figure 5: Temperature variation trend in 117 ADs in Shanxi, 2015 - 2017

Table 1: Univariate Moran's I for the outcome, independent and auxiliary variables

Variables	Univariate Moran's I	P-value
The number of CVD patients	0.287	0.001
Total population	0.268	0.001
HV	0.421	0.001
VV	0.317	0.001
Mean temperature	0.731	0.001
GDP per capita	0.257	0.001
percent of rural population	0.251	0.001
Percent of female	0.272	0.001

6.4 LISA maps

Figure 6 shows the local indicators of spatial association (LISA) maps of two predictor variables (HV and VV) in Shanxi, China. The two figures indicated that the low-low and high-high cluster patterns were more significant in HV (left) than in VV (right).

Figure 6: Local Moran's I maps of HV (left) and VV (right) in Shanxi, China

6.5 Poisson models

Table 2 shows parameter estimates of the two spatial Poisson regression without any smoothing splines. The results of the Poisson model and the hierarchical Poisson model were

consistent. The incidence rate ratio of HV and VV were both greater than one in both models, which meant that both HV and VV were associated with a higher rate of CVD hospitalization. Specifically in the Poisson model, one unit increase in HV was associated with 7.5% (95% confidence interval, CI: [7.2%-7.7%]) higher rate of CVD hospitalization, while one unit increase in VV was associated with 0.2% (95% CI: [0.1%-0.3%]) higher rate of CVD hospitalization. Similarly in the hierarchical Poisson model, one unit increase in HV was associated with 7.2% (95% CI: [6%-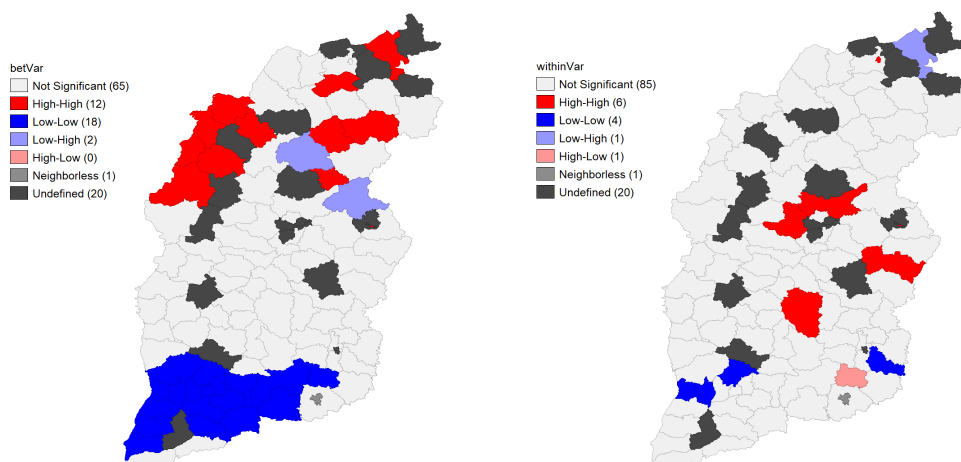8.3%]) higher rate of CVD hospitalization, while one unit increase in VV was associated with 1.5% (95% CI: [0.8%-2.3%]) higher rate of CVD hospitalization.

Table 2: Incidence rate ratio estimates of non-spatial Poisson models

	Dependent variable:	
	Poisson	Mixed-effects Poisson
	(1)	(2)
Horizontal variance	1.075***	1.072***
	(1.072,1.077)	(1.060,1.083)
Vertical variance	1.002***	1.015***
	(1.001,1.003)	(1.008,1.023)
Mean temperature	0.999***	0.998***
	(0.999,1.000)	(0.997,0.998)
Log GDP	1.224***	1.528***
	(1.220,1.228)	(1.497,1.560)
Rural	0.319***	0.349**
	(0.305,0.333)	(0.074,0.624)
Female	1.123***	1.015***
	(1.120,1.125)	(1.008,1.023)
Constant	0.00000	0.0002
	(-0.123,0.123)	(-0.417,0.418)
Log Likelihood	-148,005.600	-30,719.760
Akaike Inf. Crit.	296,025.200	61,465.530

Note: *p<0.1; **p<0.05; ***p<0.01

Since the non-parametric estimates of the smoothing splines for HV and VV did not have straightforward interpretation (Wood 2017), we did not present the parameter estimates in tables. Instead, we plotted the marginal effects of HV and VV on the rate of CVD hospitalization in Figure 7 and 8. The marginal effect of HV on the rate of CVD outcomes were consistently increasing in the two additive models. However, VV seemed to have a U-shaped effect on the rate of CVD hospitalization in non-hierarchical Poisson model but a more linear effect in the hierarchical Poisson model.
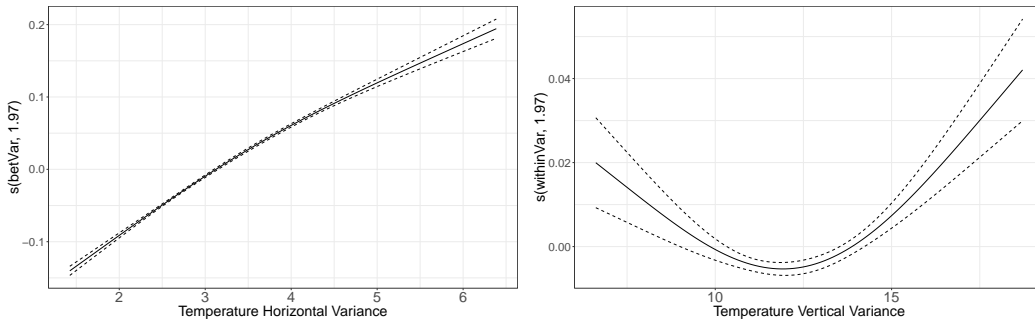
Figure 7: Marginal effects of HV (left) and VV (right) on the rate of CVD hospitalization estimated by an additive Poisson model
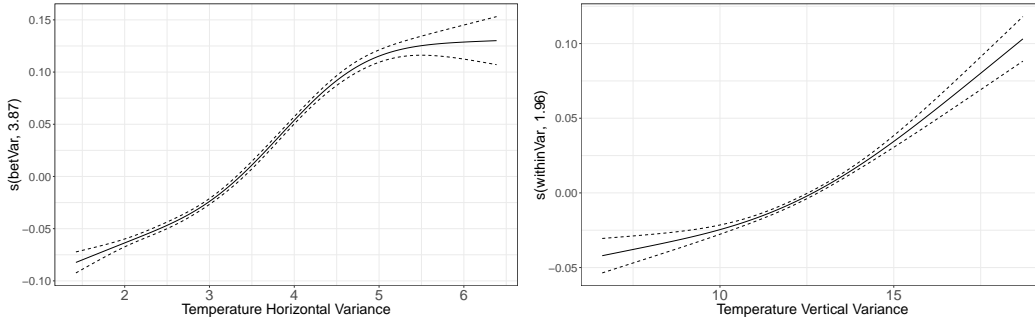
Figure 8: Marginal effects of HV (left) and VV (right) on the rate of CVD hospitalization estimated by a hierarchical additive Poisson model

Table 3 presents the incidence rate ratio (IRR) estimates of the spatial lag Poisson regression returned by `R-INLA`. After comparing Table 2 and Table 3, it was found that the difference of IRR estimates for HV and VV between hierarchical Poisson model and spatial lag Poisson model was small in this study. The IRR estimate for HV was reduced from 1.072 (95% CI:

[1.06, 1.083]) in mixed-effects Poisson model to 1.023 (95% CI: [1.011, 1.035]) in spatial lag Poisson model. As for IRR estimate for VV, it decreased from 1.015 (95% CI: [1.008, 1.023]) in mixed-effects Poisson model to 1.003 (95% CI: [0.995, 1.01]) in spatial lag Poisson model, and the estimate became insignificant after accounting for the spatial structure.

Table 3: Incidence rate ratio estimates of the spatial lag Poisson model

	Dependent variable:
Between variance	1.023***
	(1.011,1.035)
Within variance	1.003
	(0.995,1.010)
Mean temperature	1.000
	(0.998,1.002)
Log GDP	1.137***
	(1.096,1.178)
percent of the rural population	48.899***
	(48.443,49.355)
Percent of female	0.999
	(0.992,1.006)
Constant	0.0001***
	(-0.550,0.550)
Log Likelihood	-27,485.170
Akaike Inf. Crit.	55,002.340
Note:	*p<0.1; **p<0.05; ***p<0.01

7   Discussion

This study estimated the effects of two sources of temperature variation (horizontal and vertical variation) on CVD hospitalizations using traditional Poisson models, hierarchical Poisson models with and without smoothing splines, and a spatial lag Poisson model respectively. We found that horizontal temperature variation (HV), defined as temperature change cross multiple days, was associated with a significant increase in the risk of CVD hospitalization

(IRR = 1.023, 95% CI: [1.011, 1.035]) estimated by the spatial lag Poisson model based on a sample from Shanxi, China. In contrast, the association between temperature vertical variation (also widely known as the diurnal temperature range) and CVD hospitalization was less robust and highly depended on the choice of statistical models. It was significantly associated with more CVD hospitalization in non-spatial Poisson models, but the association was not significant after accounting for spatial structures.

To the best of our knowledge, this study is the first one that examines the association between temperature variation and health outcome accounting for both temperature horizontal and vertical variation, as well as spatial structures. Our results have two major contributions. First, our study indicated that previous studies on the association between DTR and health outcomes may have made type I error since they failed to account for spatial autocorrelation. Second, we found that temperature horizontal variation, instead of diurnal temperature range, may be a more robust source of the increase in CVD hospitalization.

Previous studies that reported the association between diurnal temperature range and the occurrence of CVD events almost always used traditional statistical models without accounting for spatial structure, likely due to the underdevelopment of spatial models. For example, a traditional Poisson regression was used in the study by Liang et al. (2008), a generalized linear model for temperature-matched case-crossover data by Lim, Hong, and Kim (2012), mixed-effects and generalized additive mixed Poisson models by Lim et al. (2013), a distributed lag nonlinear model by Luo et al. (2013), and a quasi-Poisson model with distributed lag non-linear model by Yang et al. (2013). These models all ignored the spatial clustering effect among geographically close neighbors, and therefore may expand the true effect by attributing the effect that could have been explained by geographic structure to observed effect (Blangiardo and Cameletti 2015). Our study supported this point since the IRR estimate of DTR (VV) was insignificant after accounting for spatial clustering in the spatial lag Poisson model, compared with the significant results in non-spatial Poisson models.

Besides, our study also found that horizontal temperature variation was a more robust risk factor for CVD hospitalization compared with DTR, after adjusting for spatial clustering.

The horizontal variation defined in this study can also be interpreted as the temperature change in a consecutive of days, which suggested a cumulative effect of temperature variation in a period is more important than temperature change within a single day.

Our study has several strengths. First, instead of deepening the current understanding of extreme temperature on human being's health outcomes, this study focuses a relatively new hypothesis on the association between temperature variation and CVD hospitalization. Second, we decomposed temperature variation into two sources, HV and VV. These two sources of temperature variation provided a more comprehensive framework for measuring temperature change, compared with previous studies that primarily focused on diurnal temperature range (VV). Third, we used a spatial lag Poisson model that accounted for the clustering effect of geographically near neighbors, which has been generally ignored in previous publications in this field.

This study should be interpreted with several limitations. First, the study was subject to potential omitted variable bias. We did not have access to several important variables such as the percent of older population and air pollution level, which were considered as important predictors of CVD hospitalization indicated by Basu et al. (2012). Second, since we aggregated hospitalized data to the AD level, we were implicitly assuming that a CVD patient who was admitted to a hospital located at a district was a resident in that district, which may not be necessarily true. Patients with severe CVD symptoms resided in remote rural areas may seek medical care in core urban cores, in which the best health care experts and high-tech medical equipment were concentrated (Cai et al. 2018). Nonetheless, very few patients will seek care in areas not within their residence area since they would have much less insurance coverage due to policy and higher traveling costs. Third, we did not compare other spatial Poisson models, such as the spatial lag Poisson model and the geographically weighted Poisson model, due to lack of available software packages. Fourth, this study was conducted at AD level, which was crude and relatively imprecise compared with census tract or block level data.

8 Conclusion

Temperature variation across multiple days is a robust risk factor for CVD hospitalization. we advise local health administration to send weather warnings to susceptible populations on upcoming drastic temperature change, which may help reduce potential cardiovascular disease hospitalization.

9 Critical Reflection

- In this study, we only used a spatial lag Poisson model without comparing the results of spatial error Poisson model or geographically weighted Poisson model. This is due to a lack of efficient estimators and computing software packages for these models (Simoes and Natário 2016; Glaser 2017).

- We set the lag term of temperature change as seven days, which has been extensively used in previous literature. This can be further explored by comparing the sensitivity of IRR estimates and their significance using different lag terms.

- Missing data on the percent of older population and air pollution can be partially solved by merging the city level data to AD level, but the data are less accurate since multiple ADs can be located within a single city. Nonetheless, the model accounting for less accurate elderly population data is better than the model without any information about these crucial variables.

References

Bakka, Haakon, Håvard Rue, Geir-Arne Fuglstad, Andrea I. Riebler, David Bolin, Janine Illian, Elias Krainski, Daniel P. Simpson, and Finn K. Lindgren. 2018. "Spatial Modelling with INLA: A Review."

Basu, Rupa, Dharshani Pearson, Brian Malig, Rachel Broadwin, and Rochelle Green. 2012. "The Effect of High Ambient Temperature on Emergency Room Visits." *Epidemiology*, 813–20.

Basu, Rupa, and Jonathan M Samet. 2002. "Relation Between Elevated Ambient Temperature and Mortality: A Review of the Epidemiologic Evidence." *Epidemiologic Reviews* 24 (2): 190–202.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. "Fitting Linear Mixed-Effects Models Using lme4." *Journal of Statistical Software* 67 (1): 1–48. https://doi.org/10.18637/jss.v067.i01.

Benziger, Catherine, Gregory Roth, and Andrew Moran. 2016. "The Global Burden of Disease Study and the Preventable Burden of NCD." *Global Heart* 11 (4): 393–97.

Blangiardo, Marta, and Michela Cameletti. 2015. *Spatial and Spatio-Temporal Bayesian Models with R-Inla.* John Wiley & Sons.

Braga, Alfésio LF, Antonella Zanobetti, and Joel Schwartz. 2002. "The Effect of Weather on Respiratory and Cardiovascular Deaths in 12 Us Cities." *Environmental Health Perspectives* 110 (9): 859–63.

Bureau of Statistics, Shanxi. 2018. "Shanxi Statistical Yearbook 2017." http://www.stats-sx.gov.cn/tjsj/tjnj/nj2017/indexch.htm.

Cai, Miao, Echu Liu, Hongbing Tao, Zhengmin Qian, Xiaojun Lin, and Zhaohui Cheng. 2018. "Does Level of Hospital Matter? A Study of Mortality of Acute Myocardial Infarction Patients in Shanxi, China." *American Journal of Medical Quality* 33 (2): 185–92.

Cheng, Jian, Zhiwei Xu, Rui Zhu, Xu Wang, Liu Jin, Jian Song, and Hong Su. 2014. "Impact of Diurnal Temperature Range on Human Health: A Systematic Review." *International Journal of Biometeorology* 58 (9): 2011–24.

Cheng, Xiaoshu, and Hai Su. 2010. "Effects of Climatic Temperature Stress on Cardiovascular Diseases." *European Journal of Internal Medicine* 21 (3): 164–67. https://doi.org/https://doi.org/10.1016/j.ejim.2010.03.001.

Glaser, Stephanie. 2017. "A Review of Spatial Econometric Models for Count Data." Hohenheim Discussion Papers in Business, Economics; Social Sciences.

Gronlund, Carina J, Antonella Zanobetti, Joel D Schwartz, Gregory A Wellenius, and Marie S O'Neill. 2014. "Heat, heat waves, and hospital admissions among the elderly in the United States, 1992–2006." *Environmental Health Perspectives* 122 (11): 1187–92.

Hu, Kejia, Yuming Guo, Xuchao Yang, Jieming Zhong, Fangrong Fei, Feng Chen, Qi Zhao, et al. 2019. "Temperature Variability and Mortality in Rural and Urban Areas in Zhejiang Province, China: An Application of a Spatiotemporal Index." *Science of the Total Environment* 647: 1044–51. https://doi.org/https://doi.org/10.1016/j.scitotenv.2018.08.095.

Jiang, Lixin, Harlan M Krumholz, Xi Li, Jing Li, and Shengshou Hu. 2015. "Achieving Best Outcomes for Patients with Cardiovascular Disease in China by Enhancing the Quality of Medical Care and Establishing a Learning Health-Care System." *The Lancet* 386 (10002): 1493–1505.

Kurmi, Om P, Julien Vaucher, Dan Xiao, Michael V Holmes, Yu Guo, Kourtney J Davis, Chen Wang, et al. 2016. "Validity of Copd Diagnoses Reported Through Nationwide Health Insurance Systems in the People's Republic of China." *International Journal of Chronic Obstructive Pulmonary Disease* 11: 419.

Li, Jing, Xi Li, Qing Wang, Shuang Hu, Yongfei Wang, Frederick A Masoudi, John A Spertus, et al. 2015. "ST-Segment Elevation Myocardial Infarction in China from 2001 to 2011 (the China Peace-Retrospective Acute Myocardial Infarction Study): A Retrospective Analysis of Hospital Data." *The Lancet* 385 (9966): 441–51.

Liang, Wen-Miin, Wen-Pin Liu, Sze-Yuan Chou, and Hsien-Wen Kuo. 2008. "Ambient Temperature and Emergency Room Admissions for Acute Coronary Syndrome in Taiwan." *International Journal of Biometeorology* 52 (3): 223–29.

Lim, Youn-Hee, Yun-Chul Hong, and Ho Kim. 2012. "Effects of Diurnal Temperature Range on Cardiovascular and Respiratory Hospital Admissions in Korea." *Science of the Total Environment* 417: 55–60.

Lim, Youn-Hee, Ho Kim, Jin Hee Kim, Sanghyuk Bae, and Yun-Chul Hong. 2013. "Effect of Diurnal Temperature Range on Cardiovascular Markers in the Elderly in Seoul, Korea." *International Journal of Biometeorology* 57 (4): 597–603.

Lin, Shao, Ming Luo, Randi J. Walker, Xiu Liu, Syni-An Hwang, and Robert Chinery. 2009. "Extreme High Temperatures and Hospital Admissions for Respiratory and Cardiovascular Diseases." *Epidemiology* 20 (5): 738–46. http://www.jstor.org/stable/25662747.

Luo, Yuan, Yonghui Zhang, Tao Liu, Shannon Rutherford, Yanjun Xu, Xiaojun Xu, Wei Wu, et al. 2013. "Lagged Effect of Diurnal Temperature Range on Mortality in a Subtropical Megacity of China." *PloS One* 8 (2): e55280.

Michelozzi, Paola, Gabriele Accetta, Manuela De Sario, Daniela Dippoliti, Claudia Marino, Michela Baccini, Annibale Biggeri, et al. 2009. "High Temperature and Hospitalizations for Cardiovascular and Respiratory Causes in 12 European Cities." *American Journal of Respiratory and Critical Care Medicine* 179 (5): 383–89. https://doi.org/10.1164/rccm.200802-217oc.

National Health and Family Planning Commission, The People's Republic of China. 2018. "China Statistical Yearbook in Health 2017." Peking Union Medical College Press.

Onozuka, Daisuke, and Akihito Hagihara. 2017. "Associations of Day-to-Day Temperature Change and Diurnal Temperature Range with Out-of-Hospital Cardiac Arrest." *European Journal of Preventive Cardiology* 24 (2): 204–12. https://doi.org/10.1177/2047487316674818.

Pan, W-H., L-A. Li, and M-J. Tsai. 1995. "Temperature Extremes and Mortality from Coronary Heart Disease and Cerebral Infarction in Elderly Chinese." *The Lancet* 345 (8946): 353–55. https://doi.org/https://doi.org/10.1016/S0140-6736(95)90341-0.

Pebesma, Edzer. 2018. "Simple Features for R: Standardized Support for Spatial Vector Data." *The R Journal* 10 (1): 439–46. https://doi.org/10.32614/RJ-2018-009.

Perkins, SE, LV Alexander, and JR Nairn. 2012. "Increasing Frequency, Intensity and Duration of Observed Global Heatwaves and Warm Spells." *Geophysical Research Letters* 39 (20).

Phung, Dung, Phong K Thai, Yuming Guo, Lidia Morawska, Shannon Rutherford, and Cordia Chu. 2016. "Ambient Temperature and Risk of Cardiovascular Hospitalization: An Updated Systematic Review and Meta-Analysis." *Science of the Total Environment* 550: 1084–1102.

Prüss-Üstün, Annette, and Maria Neira. 2016. *Preventing Disease Through Healthy Environments: A Global Assessment of the Burden of Disease from Environmental Risks*. World Health Organization.

Qiu, Hong, Ignatius Tak-sun Yu, Lap Ah Tse, Linwei Tian, Xiaorong Wang, and Tze Wai Wong. 2013. "Is Greater Temperature Change Within a Day Associated with Increased Emergency Hospital Admissions for Heart Failure?" *Circulation: Heart Failure* 6 (5): 930–35.

R Core Team. 2018. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Rue, Håvard, Andrea I. Riebler, Sigrunn H. Sørbye, Janine B. Illian, Daniel P. Simpson, and Finn K. Lindgren. 2017. "Bayesian Computing with INLA: A Review."

Ryti, Niilo RI, Yuming Guo, and Jouni JK Jaakkola. 2015. "Global Association of Cold Spells and Adverse Health Effects: A Systematic Review and Meta-Analysis." *Environmental Health Perspectives* 124 (1): 12–22.

Sartini, Claudio, Sarah JE Barry, Peter H Whincup, S Goya Wannamethee, Gordon DO Lowe, Barbara J Jefferis, Lucy Lennon, et al. 2017. "Relationship Between Outdoor Temperature and Cardiovascular Disease Risk Factors in Older People." *European Journal of Preventive Cardiology* 24 (4): 349–56.

Shinkawa, Atsushi, Kazuo Ueda, Yutaka Hasuo, Yutaka Kiyohara, and Masatoshi Fujishima. 1990. "Seasonal Variation in Stroke Incidence in Hisayama, Japan." *Stroke* 21 (9): 1262–7.

Simoes, Paula, and Isabel Natário. 2016. "Spatial Econometric Approaches for Count Data: An Overview and New Directions." *IntJ Soc Behav Educ Econ Bus Ind Eng* 10 (1): 348–56.

The World Health Organization. 2018. "The Top 10 Causes of Death." http://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death.

Wang, Min-zhen, Shan Zheng, Shi-lin He, Bei Li, Huai-jin Teng, Shi-gong Wang, Ling Yin, Ke-zheng Shang, and Tan-shi Li. 2013. "The Association Between Diurnal Temperature Range and Emergency Room Admissions for Cardiovascular, Respiratory, Digestive and Genitourinary Disease Among the Elderly: A Time Series Study." *Science of the Total Environment* 456: 370–75.

Watts, Nick, W Neil Adger, Sonja Ayeb-Karlsson, Yuqi Bai, Peter Byass, Diarmid Campbell-Lendrum, Tim Colbourn, et al. 2017. "The Lancet Countdown: Tracking Progress on Health and Climate Change." *The Lancet* 389 (10074): 1151–64.

Wickham, Hadley. 2017. *Tidyverse: Easily Install and Load the 'Tidyverse'*. https://CRAN.R-project.org/package=tidyverse.

Wikipedia contributors. 2019. "Shanxi — Wikipedia, the Free Encyclopedia." https://en.wikipedia.org/w/index.php?title=Shanxi&oldid=895434579.

Wood, S.N. 2017. *Generalized Additive Models: An Introduction with R*. 2nd ed. Chapman; Hall/CRC.

Yang, Jun, Hua-Zhang Liu, Chun-Quan Ou, Guo-Zhen Lin, Qin Zhou, Gi-Chuan Shen, Ping-Yan Chen, and Yuming Guo. 2013. "Global Climate Change: Impact of Diurnal Temperature Range on Mortality in Guangzhou, China." *Environmental Pollution* 175: 131–36.

Yang, Jun, Peng Yin, Maigeng Zhou, Chun-Quan Ou, Yuming Guo, Antonio Gasparrini, Yunning Liu, et al. 2015. "Cardiovascular Mortality Risk Attributable to Ambient Temperature in China." *Heart* 101 (24): 1966–72.

Yang, Jun, Maigeng Zhou, Mengmeng Li, Peng Yin, Boguang Wang, Eva Pilot, Yunning Liu, et al. 2018. "Diurnal Temperature Range in Relation to Death from Stroke in China." *Environmental Research* 164: 669–75. https://doi.org/https://doi.org/10.1016/j.envres.2018.03.036.

Yang, Ling, Liming Li, Sarah Lewington, Yu Guo, Paul Sherliker, Zheng Bian, Rory Collins, et al. 2015. "Outdoor Temperature, Blood Pressure, and Cardiovascular Disease Mortality

Among 23 000 Individuals with Diagnosed Cardiovascular Diseases from China." *European Heart Journal* 36 (19): 1178–85.

Ye, Xiaofang, Rodney Wolff, Weiwei Yu, Pavla Vaneckova, Xiaochuan Pan, and Shilu Tong. 2011. "Ambient Temperature and Morbidity: A Review of Epidemiological Evidence." *Environmental Health Perspectives* 120 (1): 19–28.

Zheng, Shan, Minzhen Wang, Bei Li, Shigong Wang, Shilin He, Ling Yin, Kezheng Shang, and Tanshi Li. 2016. "Gender, Age and Season as Modifiers of the Effects of Diurnal Temperature Range on Emergency Room Admissions for Cause-Specific Cardiovascular Disease Among the Elderly in Beijing." *International Journal of Environmental Research and Public Health* 13 (5). https://doi.org/10.3390/ijerph13050447.

A R code for data visualization

A.1 Distribution of Shanxi hospitals

```
library("scales")


load("data/hloca.Rdata")
names(hloca)[names(hloca) == "longtitude"] = "longitude"


ggplot() +
  geom_sf(data = sx, fill = "white", color = "grey") +
  geom_point(data = hloca,
             aes(x=longitude, y=latitude, size=N),
             shape=21, colour = "blue",
             fill="blue", alpha = 0.5) +
  scale_size_continuous(labels = comma, range = c(1, 10),
                        name = "The number\nof patients")+
  theme_bw() + theme(axis.title.x=element_blank(),
                     axis.title.y=element_blank())


ggsave("figs/hospitals.pdf", height=7, width=6)
```

A.2 Patient choropleth map

```
# Sys.setlocale(category ="LC_ALL", locale ="zh_cn.utf-8")
sx1 <- sf::st_read("data/shp/sx_border.shp") %>%
  left_join(distpat, by = c("ADMINCODE" = "dist_old")) %>%
  mutate(city = tolower(city), county = tolower(county)) %>%
  mutate(
    pat_cat = NA,
    pat_cat = ifelse(Npat < 2150, "<2150", pat_cat),
```

```r
    pat_cat = ifelse(Npat >= 2150 & Npat < 4400, "2150-4400", pat_cat),

    pat_cat = ifelse(Npat >= 4400 & Npat < 7700, "4400-7700", pat_cat),

    pat_cat = ifelse(Npat >= 7700 & Npat < 13000, "7700-13000", pat_cat),

    pat_cat = ifelse(Npat >= 13000, ">13000", pat_cat)

  ) %>%

  mutate(pat_cat = factor(pat_cat,

    levels =

      c(">13000", "7700-13000", "4400-7700", "2150-4400", "<2150")

  ))


ggplot() +

  geom_sf(

    data = sx1, size = 0.1,

    color = "grey", aes(fill = pat_cat)

  ) + # #404040

  scale_fill_brewer(palette = "Reds", direction = -1) +

  coord_sf(crs = "+proj=aeqd +lat_0=37 +lon_0=104") +

  labs(fill = "The number of \nCVD patients") +

  theme_bw()
# + geom_sf_text(data = sx1, aes(label = PYNAME), colour = "blue")


ggsave("figs/choropleth_patients.pdf", height = 7, width = 6)
```

A.3   The number of patients by month

```r
CVDsx <- fread("data/CVDsx.csv")

hosp_dist <- fread("data/hosp_district.csv") %>%

  set_names(c("hospid", "district", "city", "county"))


res_date <- c(201501:201512, 201601:201612, 201701:201712) %>%
```

```r
  as.character()
zdist <- CVDsx[, .N, c("hospid", "yearmonth")][
  hosp_dist,
  on = "hospid"
][
  , .(N = sum(N)),
  by = c("district", "city", "county", "yearmonth")
][
  , yearmonth := factor(yearmonth, levels = res_date)
]
zcity <- CVDsx[, .N, c("hospid", "yearmonth")][
  hosp_dist,
  on = "hospid"
][
  , .(N = sum(N)),
  by = c("city", "yearmonth")
][
  , yearmonth := factor(yearmonth, levels = res_date)
]


pday <- CVDsx[, yearmonth := factor(
  yearmonth,
  levels = res_date
)][, .N, "yearmonth"]

pday %>%
  ggplot(aes(yearmonth, N)) +
  geom_bar(stat = "identity", fill = "#1e32cc") + theme_bw() +
  scale_y_continuous(labels = scales::comma,
```

```r
                     name = "The number of CVD inpatients") +
  xlab("Year and month") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))


ggsave("figs/Npat_by_month.pdf", width = 10, height = 6.18)
```

A.4   The maximum and minimum temperature by month in 117 ADs

```r
require("pinyin")
mypy <- pydic(method = "toneless", dic = "pinyin2")
load("data/weather.Rdata")
res_date <- c(201501:201512, 201601:201612, 201701:201712) %>%
  as.character()


w <- weather %>%
  select(1, 2,
    mdate = ymd, max_tem = bWendu,
    min_tem = yWendu, aqi, aqiLevel
  ) %>%
  mutate(myear = as.integer(substr(mdate, 1, 4))) %>%
  filter(myear >= 2015) %>%
  mutate(
    mdate = lubridate::ymd(mdate),
    max_tem = gsub("([0-9]+).*$", "\\1", max_tem) %>% as.numeric(),
    min_tem = gsub("([0-9]+).*$", "\\1", min_tem) %>% as.numeric(),
    yearmonth = factor(paste0(
      substr(mdate, 1, 4),
      substr(mdate, 6, 7)
    )),
    city = py(city, dic = mypy, sep = ""),
```

```r
    county = py(county, dic = mypy, sep = ""),
    tem_diff = max_tem - min_tem,
    tem_mean = (max_tem + min_tem) / 2
  ) %>%
  mutate(
    city = gsub("[[:digit:]]+", "", city),
    county = gsub("[[:digit:]]+", "", county)
  ) %>%
  arrange(city, county, mdate)


w1 <- w %>%
  mutate(
    county = gsub("changzhi", "shangdang", county),
    county = gsub("changzi", "zhangzi", county),
    county = gsub("datong", "pingcheng", county),
    county = gsub("datongxian", "yungang", county),
    county = gsub("jincheng", "chengqu", county),
    county = gsub("hongdong", "hongtong", county),
    county = gsub("yaodou", "yaodu", county),
    county = gsub("jiancaopingqu", "jiancaoping", county),
    county = gsub("xiaodianqu", "xiaodian", county),
    county = gsub("fanzhi", "fanshi", county),
    county = gsub("jingyue", "jingle", county),
    county = gsub("wutaishan", "wutai", county),
    county = gsub("yangquan", "chengqu", county)
  )
w1_re <- w %>%
  filter(county %in% c("changzhi", "datong", "jiaoqu")) %>%
  mutate(
    county = gsub("changzhi", "luzhou", county),
```

```r
      county = gsub("datong", "yungang", county),
      county = gsub("jiaoqu", "kuangqu", county)
  )


w1 <- w1 %>%
  rbind(w1_re)



c2f <- function(x) return((x * 9 / 5) + 32)

w1 %>%
  group_by(county, yearmonth) %>%
  summarise(
    min_tem = mean(min_tem),
    max_tem = mean(max_tem)
  ) %>%
  ungroup() %>%
  mutate(
    max_tem = c2f(max_tem),
    min_tem = c2f(min_tem)
  ) %>%
  ggplot() +
  geom_line(aes(yearmonth, max_tem,
    group = county,
    color = "max"
  ), alpha = 0.3) +
  geom_line(aes(yearmonth, min_tem,
    group = county,
    color = "min"
  ), alpha = 0.3) +
```

```r
  scale_colour_manual(
    name = "Line Color",
    values = c(max = "red", min = "blue")
  ) +
  theme_bw() + ylab("Temperature in Fahrenheits") +
  xlab("Year and month") +
  theme(
    legend.justification = c(1, 1), legend.position = c(0.9, 0.12),
    legend.background = element_rect(fill = alpha("white", 0.4)),
    legend.direction = "horizontal",
    axis.text.x = element_text(angle = 45, hjust = 1)
  ) +
  guides(color = guide_legend(
    title = "Temperature type",
    override.aes = list(alpha = 1, size = 2)
  ))
ggsave("figs/temperature_in_groups.pdf", width = 10, height = 6.18)
```

A.5   Two sources of temperature variation by month in 117 ADs

```r
lagterm <- 7
aw <- w %>%
  group_by(city, county) %>%
  mutate(
    lagtem_mean = lag(tem_mean, lagterm),
    lagtem_diff = lag(tem_diff, lagterm)
  ) %>%
  ungroup() %>%
  group_by(city, county, yearmonth) %>%
  summarise(
```

```r
    betweenVar = sd(lagtem_mean, na.rm = TRUE),
    withinVar = mean(lagtem_diff, na.rm = TRUE)
  ) %>%
  ungroup()


aw %>%
  gather("var_type", "value", -city, -county, -yearmonth) %>%
  mutate(group0 = paste0(city, county, var_type)) %>%
  ggplot(aes(yearmonth, value, group = group0, color = var_type)) +
  geom_line(alpha = 0.2) + theme_bw() +
  xlab("Year and month") + ylab("Temperature variation") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  theme(
    legend.justification = c(1, 1), legend.position = c(0.7, 0.99),
    legend.background = element_rect(fill = alpha("white", 0.4)),
    legend.direction = "horizontal"
  ) +
  guides(color = guide_legend(
    title = "Type of variation",
    override.aes = list(alpha = 1, size = 2)
  )) +
  scale_color_manual(
    values = c("red", "blue"),
    labels = c("Horizontal Variation", "Vertical Variation")
  )

ggsave("figs/two_tem_variation.pdf", width = 10, height = 6.18)
```

B   R code for statistical models

B.1   Poisson model

```r
fit1 <- glm(Npat ~ betVar + withinVar + tmean +
                log(gdp_cap) + p_rural + p_female,
            data = sx_poi, offset = log(pop100t),
            family = poisson(link = "log")
)
summary(fit1)$coefficients
save(fit1, file = "data/fit1.Rdata")
```

B.2   Hierarchical Poisson model

```r
require(lme4)
fit2 <- glmer(Npat ~ betVar + withinVar + tmean +
                log(gdp_cap) + p_rural + p_female +
                (1 + betVar + withinVar|citycounty),
            offset = log(pop100t), data = sx_poi,
            family = poisson(link = "log"),nAGQ=0)
summary(fit2)$coefficients
save(fit2, file = "data/fit2.Rdata")
```

B.3   Spatial lag Poisson model

```r
require(INLA)
poi_inla = inla(fpp, family = "poisson", data = sx_poi,
                control.compute = list(dic = TRUE),
                control.inla = list(tolerance = 1e-20, h = 1e-08),
                control.predictor = list(compute = TRUE))
```