Environmental Research 138 (2015) 439–452

Contents lists available at ScienceDirect

Environmental Research

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

Geographic dimensions of heat-related mortality in seven U.S. cities

David M. Hondula [a,b,*], Robert E. Davis [c], Michael V. Saha [c], Carleigh R. Wegner [c], Lindsay M. Veazey [c]

[a] Center for Policy Informatics, School of Public Affairs, Arizona State University, Phoenix, AZ 85004, USA
[b] School of Geographical Sciences and Urban Planning, Arizona State University, Tempe, AZ 85287, USA
[c] Department of Environmental Sciences, University of Virginia, Charlottesville, VA 22904, USA

ARTICLE INFO

Article history:
Received 15 September 2014
Received in revised form
7 February 2015
Accepted 26 February 2015
Available online 17 March 2015

Keywords:
Heat
Mortality
Spatial
Vulnerability
Urban

ABSTRACT

Spatially targeted interventions may help protect the public when extreme heat occurs. Health outcome data are increasingly being used to map intra-urban variability in heat-health risks, but there has been little effort to compare patterns and risk factors between cities. We sought to identify places within large metropolitan areas where the mortality rate is highest on hot summer days and determine if characteristics of high-risk areas are consistent from one city to another. A Poisson regression model was adapted to quantify temperature–mortality relationships at the postal code scale based on 2.1 million records of daily all-cause mortality counts from seven U.S. cities. Multivariate spatial regression models were then used to determine the demographic and environmental variables most closely associated with intra-city variability in risk.

Significant mortality increases on extreme heat days were confined to 12–44% of postal codes comprising each city. Places with greater risk had more developed land, young, elderly, and minority residents, and lower income and educational attainment, but the key explanatory variables varied from one city to another. Regression models accounted for 14–34% of the spatial variability in heat-related mortality. The results emphasize the need for public health plans for heat to be locally tailored and not assume that pre-identified vulnerability indicators are universally applicable. As known risk factors accounted for no more than one third of the spatial variability in heat–health outcomes, consideration of health outcome data is important in efforts to identify and protect residents of the places where the heat-related health risks are the highest.

1. Introduction

Forecasts of more severe and frequent heat waves in the future have captured the attention of public health officials and researchers. Extreme heat already ranks as a leading weather-related cause of death in the United States (NCHS, 2014), and the possibility that the related public health burden will increase in the future has motivated a range of stakeholders to pursue new strategies to protect citizens (Lowe et al., 2011; Yardley et al., 2011).

Much of the existing knowledge on the relationship between high temperatures (and humidity) and human health is derived from studies using aggregated data representing the entire populations of large cities. The discovery that the relationship varies from one city to another (e.g., Kalkstein and Davis, 1989; Curriero et al., 2002; Pascal et al., 2006) was instrumental in advancing the city-specific heat–health warning systems currently operating across the globe (Sheridan and Kalkstein, 2004; Hondula et al., 2013b). An underlying premise of these systems is that the population's sensitivity to high temperatures varies spatially, and thus the threshold temperature for activating warning systems and deploying resources for interventions should also vary. Thus, information about geographic variability in the response *between* cities is already motivating spatially targeted intervention and mitigation activities.

There is ample evidence supporting the notion that the response to high temperatures might also be spatially variable *within* cities. Important determinants of this intra-city variability in heat-related risk include vulnerability related to demographic, social, and economic characteristics (e.g., young and old age (Green et al., 2001; Koppe et al., 2004; Gosling et al., 2009), lower socioeconomic status (Kilbourne et al., 1982; Naughton et al., 2002; Harlan et al., 2013), and social isolation (Semenza et al., 1996; Kalkstein and Sheridan, 2007). There may also be fine-scale differences in exposure to high temperatures arising from urbanization effects (Smargiassi et al., 2009; Laaidi et al., 2012). When

* Corresponding author at: 411 N. Central Avenue, Suite 400, Center for Policy Informatics, Arizona State University, Phoenix, AZ, USA.
E-mail address: david.hondula@asu.edu (D.M. Hondula).

http://dx.doi.org/10.1016/j.envres.2015.02.033
0013-9351/

examined collectively, these variables and others can lead to a spatially complex risk-scape for heat-related mortality. The accessibility of spatially referenced information about heat vulnerability factors has motivated an increasing number of researchers to derive various heat vulnerability indices from this suite of variables and examine intra-urban variability in these indices (e.g., Reid et al., 2009; Chow et al., 2012; Johnson et al., 2012; Wolf and McGregor, 2013). Public health agencies have readily integrated heat vulnerability mapping into their own preparedness documents and climate adaptation plans (e.g., MPHI, 2011; CDC, 2012; SFDPH, 2012; Loughnan et al., 2013). However, research examining how heat-health outcomes compare spatially with anticipated vulnerability factors is sparse (e.g., Uejio et al., 2011; Reid et al., 2012; Harlan et al., 2013; Hondula and Barnett, 2014), and limited accessibility to long-term, consistent health outcome data has inhibited robust multi-city assessments. As a result, our understanding of the utility of heat vulnerability indices to identify places in the greatest need of targeted intervention strategies may be limited.

We use multidecadal, geographically referenced medical records to pursue two objectives: (1) to identify locations within cities where the mortality rate is highest during extreme heat events, and (2) to understand the socioeconomic and environmental factors associated with high-risk zones and their applicability from one place to another. Such information can facilitate more targeted and effective intervention measures by helping health and emergency management officials determine where and how they should allocate public resources to combat negative consequences of extreme heat events (Ebi and Schmier, 2005).

2. Methods

2.1. Data sources

Daily mortality records including the postal code of residence of the decedent were obtained for seven major metropolitan areas in the United States (Atlanta, Georgia; Boston, Massachusetts; Minneapolis-St. Paul, Minnesota; Philadelphia, Pennsylvania; Phoenix, Arizona; Seattle, Washington; St. Louis, Missouri) that span multiple climate zones. Data were sourced from the respective state departments of health. On average, 22 years of data were available for each city; the period of records varied slightly based on data availability (Table 1). In total, 2,117,584 cases were examined. There were no periods of missing data. Spatial boundaries for each city were chosen to include the populated metropolitan core and immediate surrounds. The number of postal codes included per city ranged from 63 to 101 (see Supplemental materials). Institutional Review Board approval was not required for this research because the data were de-identified prior to receipt from the providing offices.

Hourly meteorological data for each city were obtained from the archives of the United States National Climatic Data Center (http://www.ncdc.noaa.gov). The data were selected from the first-order weather station closest to each study city. The air temperature, dew point temperature, and wind speed time series had few missing values ($<1\%$ of all observations for each station). There is a lack of consensus in the literature regarding the optimal exposure variable to use when examining warm-season temperature–mortality relationships (Barnett et al., 2010; Vaneckova et al., 2011). We calculated the daily maximum afternoon temperature, in which “afternoon” was defined as the five-hour window centered on the hour of average maximum temperature. We also calculated the daily afternoon maximum apparent temperature, a variable that combines the effects of temperature, humidity and wind, following a parameterization of the Steadman model (Steadman, 1984; Hondula et al., 2012).

Data representing social and demographic indicators (Table 2) were obtained from the United States Census Bureau from the year 2000 census. The data were downloaded for year 2000 Zip Code Tabulation Areas (ZCTAs, henceforth “postal codes”) from the National Historical Geographic Information System portal (https://www.nhgis.org) (Minnesota Population Center, 2011). Geographic boundary files for the postal codes were obtained from the same source. Land cover information was obtained from the National Land Cover Database (NLCD) through the Multi-Resolution Land Characteristics Consortium (http://www.mrlc.gov/index.php) (Fry et al., 2011). The NLCD includes 39 different classification types for 80 m^2 pixels spanning the United States, including three separate types representative of the urban environment (low, medium, and high-intensity development). We calculated the percentage of each land cover type within each postal code using the Zonal Statistics tool in ESRI ArcMap version 10.0. The non-urban NLCD types present in the study cities were aggregated into broader categories (Table 2). Information about land surface temperatures (e.g., Smargiassi et al., 2009; Hondula et al., 2012; Johnson et al., 2012) was not included in the models because of concerns regarding the representativeness of specific images taken on discrete days and times for assessing intra-urban variability in thermal conditions that emerged as we processed satellite data for some of the study cities.

2.2. Postal code temperature–mortality estimates

Postal code level heat-related mortality estimates are derived from a multi-stage statistical procedure that accounts for seasonality and long-term time trends following Hondula et al. (2013a). In the first stage, the time and temperature components of the

Table 1
Descriptive statistics of the data used for this study from each of the six cities where a statistical relationship between summer temperature and mortality was found. The table also includes information about the city-specific temperature–mortality models including the relative risk (RR) predicted by the mortality at the temperature threshold (T^*), and the average RR when the threshold is exceeded.

	Boston	Minneapolis	Philadelphia	Phoenix	Seattle	St. Louis
Descriptive Statistics						
Period of record	1987 – 2007	1992 – 2008	1983 – 2008	1989 – 2007	1988 – 2008	1980 – 2008
Average daily mortality	35.75	37.58	43.16	52.38	29.68	37.48
Average summer max. temp.	25.50	26.40	29.06	39.92	22.66	30.31
City-wide Temperature Model						
Threshold temperature (T^*)	27.90	30.00	30.70	42.50	25.90	34.10
Model-predicted RR at T^*	1.017	1.022	1.015	1.019	1.021	1.016
% of summer days T^* exceeded	26.5%	17.9%	27.9%	14.8%	22.0%	13.0%
Average RR when T^* exceeded	1.059	1.042	1.066	1.037	1.062	1.029

Table 2
Independent variables included in the study as potential factors related to spatial variability in heat-related mortality at the postal code scale. The data sources are the U.S. Census Bureau via the National Historical Geographical Information Systems (NHGIS) (Minnesota Population Center, 2011) and the National Land Cover Database (NLCD) (Fry et al., 2011).

Variable (by postal code)	Source
Percent of residents over age 65	NHGIS/U.S. Census
Percent of residents under age 5	NHGIS/U.S. Census
Percent of residents over age 25 without a high school diploma	NHGIS/U.S. Census
Percent of households with one resident	NHGIS/U.S. Census
Percent of households with one resident over age 65	NHGIS/U.S. Census
Percent of residents with public assistance for disability	NHGIS/U.S. Census
Percent of residents living in poverty	NHGIS/U.S. Census
Percent of residents who are white	NHGIS/U.S. Census
Percent of residents who are black	NHGIS/U.S. Census
Percent of residents who are American Indian	NHGIS/U.S. Census
Percent of residents who are Asian	NHGIS/U.S. Census
Percent of residents who are Pacific Islander	NHGIS/U.S. Census
Percent of residents of another race	NHGIS/U.S. Census
Percent of residents of two or more races	NHGIS/U.S. Census
Mean per capita income	NHGIS/U.S. Census
Median housing value	NHGIS/U.S. Census
Percent of dwellings built 1940 or earlier	NHGIS/U.S. Census
Percent of dwellings built 1970 or earlier	NHGIS/U.S. Census
Percent land area with open space	NLCD
Percent land area with low-intensity development	NLCD
Percent land area with medium-intensity development	NLCD
Percent land area with high-intensity development	NLCD
Percent land area with water or wetlands	NLCD
Percent land area with forest	NLCD
Percent land area with grass or crops	NLCD

time series aggregated across all postal codes within each city are estimated with a generalized additive model (Wood, 2006). The temperature–mortality relative risk curves are shown in Fig. 1; temporal components of the models for each city are available in Supplemental materials. A threshold temperature is determined from each city's aggregate (non-spatial) model by determining the lowest temperature at which mortality is significantly greater than what is expected for mean summer conditions (see Supplemental materials for additional details).

In the next stage of the model, seasonality and long-term time trends are removed from the postal code level data. As the low daily mortality counts within each sample do not permit reliable estimation of these effects by directly modeling them analogously to the city aggregate models, the shape of the seasonality curve within each year is obtained from the city aggregate model and scaled and shifted to match each postal code-year's average summer mortality count. The original postal code daily mortality counts are divided by the resultant estimated baseline mortality time series to yield a daily mortality ratio, which serves as the primary variable for analysis henceforth. For the remainder of the manuscript, the term "heat-related mortality" specifically refers to excess all-cause deaths that occur on days with high temperatures, and heat-related mortality will be reported as a risk ratio.

Finally, the average mortality ratio (M^*) on all days June–September exceeding the city-specific threshold temperature is calculated for each postal code, and a randomization test is used to identify those postal codes where average mortality on above-threshold days statistically differs (with 95% confidence) from what would be expected under normal summer conditions (Sheridan and Dolney, 2003). If there was no statistical difference between the baseline mortality rate and that observed on above-threshold days, and above-threshold day mortality at the postal code scale was spatially independent and randomly distributed, one would expect 5% of postal codes to have "statistically significantly high" mortality on above-threshold days because of making multiple comparisons. Details of this procedure for use in spatial units with low daily mortality counts are can be found in Hondula et al. (2013a).

2.3. Spatial regression

Potential associations between spatial variability in heat-related mortality rates and socioeconomic and environmental factors were explored with multivariate spatial regression models. The dependent variable was the ratio of M^* to the statistical significance threshold from the randomization test (henceforth, T). This ratio was used instead of M^* because the variance in the mortality ratio time series markedly differs from one postal code to another, making it difficult to draw meaningful comparisons from M^* alone. T provides a standardized estimate for each postal code regarding how exceptional mortality is on above-threshold days for that specific location. If mortality data for each postal code were normally distributed, the dependent variable for the analysis would be the t-statistic comparing each postal code's mean mortality to a certain baseline value. Here, T is analogous but for the randomization test accounting for non-normality.

Postal codes with low T values (classified as outliers based on standard deviation criteria and visual inspection) were excluded from the regression to avoid the likelihood of isolated leverage points having unduly large impact on the overall analysis. A majority of postal codes removed had very low populations and mortality counts. No more than five postal codes were excluded from the analysis in any given city.

All subsets linear regression was used to select variables for each city that are associated with spatial variability in heat-related mortality. An exhaustive set of all possible multivariate linear regression models was generated for each city with 1–10 independent variables included. The optimal model was selected from all possible subsets using Schwarz's information criterion (BIC) (Schwarz, 1978). The model with the lowest BIC was then examined for spatial autocorrelation in the residuals (Moran's I), collinearity (Jarque–Bera test), and heteroskedasticity (Koenker–Bassett test). Where necessary, this procedure was iteratively repeated for each city as variables were excluded based on non-normality heteroskedasticity, and/or collinearity with other terms included in the model. We also performed linear regression using unrotated principal components of the original pool of explanatory variables. Components were identified separately for each city, and those with eigenvalues greater than one were saved as new independent variables. We then generated a multiple regression model using the principal components by including all components with partial significance values less than 0.05.

Finally, we examined data from all cities simultaneously by merging all postal code level data into one single data set. The same procedures as listed above for each city were followed to generate multiple regression models using the original pool of explanatory variables and principal components. Here, the principal components analysis was completed using all cities' data combined.

The regression models were generated using the *leaps* package in RStudio version 0.96.304 (RStudio, 2012) and final modeling, diagnostics, and spatial corrections were performed with GeoDa version 1.4.0 (Anselin et al., 2006). Principal components analysis was completed with IBM SPSS Statistics version 20.0.

3. Results

3.1. Intra-city heat-related mortality

A statistically significant positive association between high temperatures and all-cause mortality, controlling for time

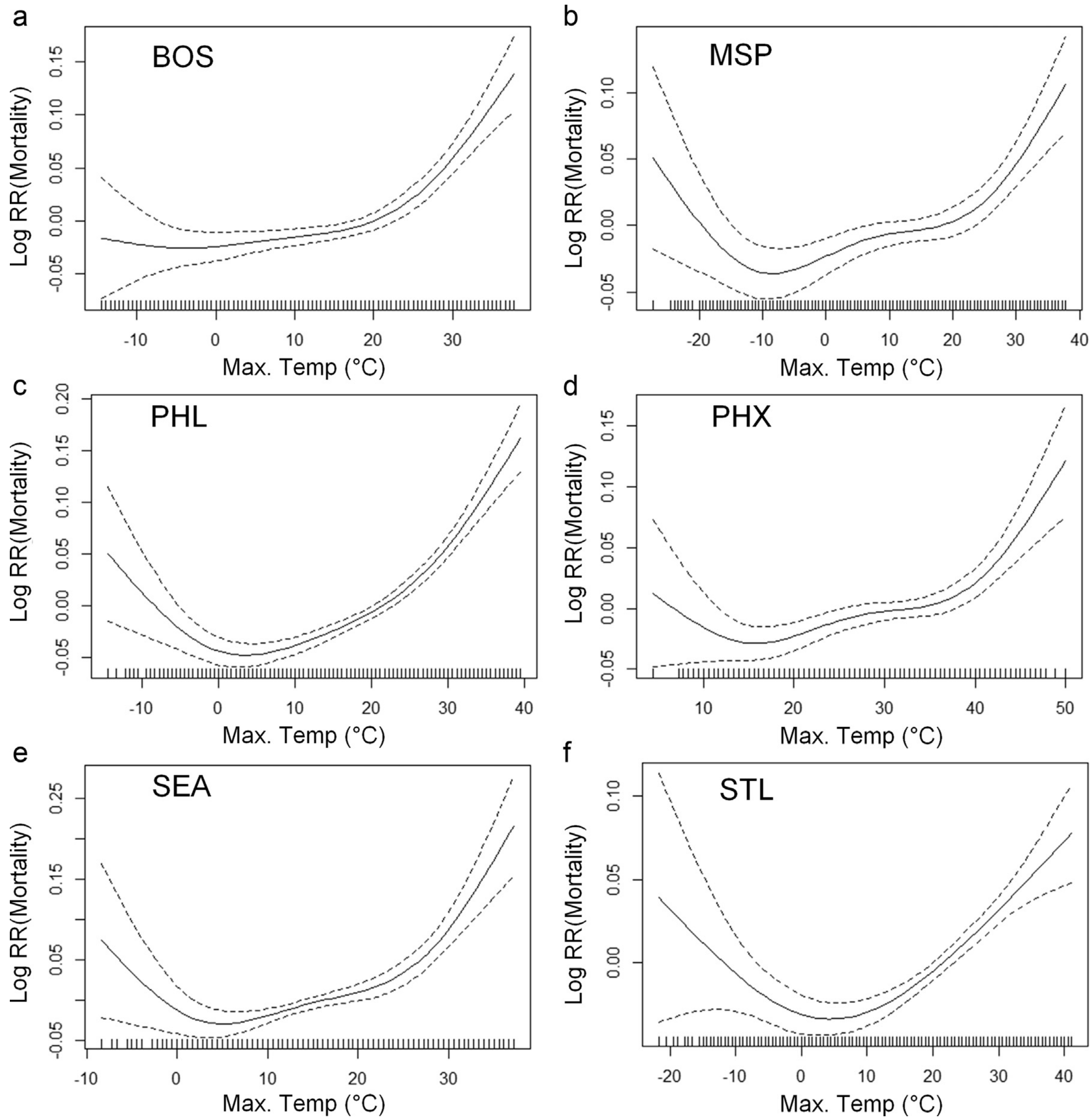

Fig. 1. (a–f). The modeled temperature–mortality relationship for each of the study cities based on historical data. The solid line is the model estimate and the dashed lines represent the confidence intervals. The tick marks on the horizontal axis show the distribution of temperature observations for each city.

confounders, was evident in six of the seven study cities (Table 1). No threshold temperature was evident in Atlanta, so Atlanta is excluded from the remainder of the analysis. Threshold temperatures for statistically significant increases in heat-related mortality varied from 1.6 °C (Philadelphia) to 3.8 °C (St. Louis) above the summer mean temperature. Threshold temperatures were exceeded on 13.0–27.9% of summer days during each city's study period. Model-predicted relative risks at the threshold temperature were largely consistent across cities, varying from 1.015 (Philadelphia) to 1.020 (Minneapolis). Larger inter-city differences were evident in the average mortality rate on days when the threshold temperature was exceeded, ranging from +2.9% in St. Louis to +6.6% in Philadelphia. Results were similar for apparent temperature, and as air temperature is simpler to measure and report, it is the exposure variable chosen for the remainder of the analysis (Barnett et al., 2010).

Significant intra-city spatial variation in mortality rates on above-threshold days was evident in each of the six cities examined (Fig. 2a–f). Intra-city differences in mortality rates exceeded inter-city differences. Significant increases over baseline summer mortality rates were present for 15 of 60 examined postal codes in Boston (25%), 13 of 100 in Minneapolis (13%), 20 of 45 in Philadelphia (44.4%), 12 of 93 in Phoenix (12.9%), 11 of 60 in Seattle (18.3%), and 7 of 60 in St. Louis (11.7%).

3.2. Spatial regression with original variables

From the original set of 25 demographic and environmental variables included as potential predictors of spatial variability in heat-related mortality rates, eight were included in multiple regression models across the six cities (Table 3). The number of variables used for each city varied from one to four. All models and individual terms were statistically significant and all models but one (Phoenix) passed diagnostic tests for lack of spatial autocorrelation in residuals. Two other diagnostic tests did not meet statistical criteria (residual heteroskedasticity in Minneapolis and residual normality in St. Louis), but in both cases, this was related to individual outlier residuals that did not substantively impact the regression. The models accounted for 14.1–33.5% of the spatial variability in heat-related mortality.

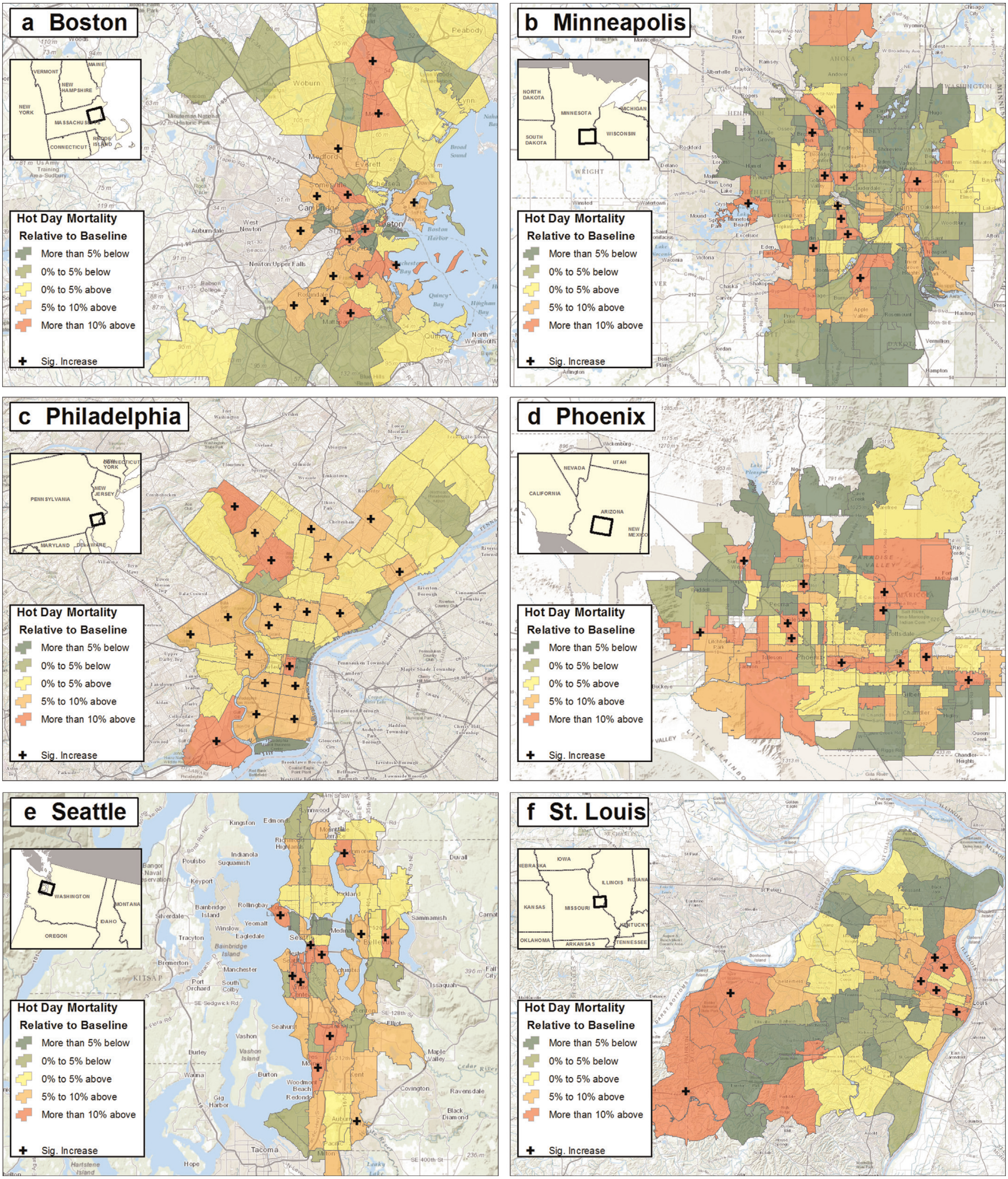

Fig. 2. (a–f). Maps showing intra-city variability in heat-related risk. The spatial units shown on the map are year 2000 Zip Code Tabulation Areas (postal codes). Each postal code is shaded according to the percent change in mortality on days that exceed a city-specific temperature threshold relative to a baseline summer mortality rate. Those postal codes where the mortality rate on threshold-exceeding days is statistically significantly greater than the baseline are identified with a plus (+) sign.

The specific set of independent variables included in the regression models varied from one city to another. Demographic variables associated with higher risk included lower per capita income (in Boston), higher percentages of elderly residents (in Minneapolis, Philadelphia, and Seattle), higher percentages of residents of Asian heritage (in Philadelphia), higher percentages of children under age five (in Philadelphia), lower percentages of white residents (in Philadelphia), lower median housing values (in

Table 3
Details from multivariate linear regression models generated for each city relating intra-urban variability in heat-related risk to demographic and environmental factors. The model summary portion of the table includes several diagnostic tests for residual normality, heteroskedasticity, and autocorrelation.

	Boston	Minneapolis	Philadelphia	Phoenix	Seattle	St. Louis
Model summary						
N	60	100	45	92	60	60
F	15.867	18.378	4.528	19.182*	6.300	10.669
p	< 0.01	< 0.01	< 0.01	< 0.01*	< 0.01	< 0.01
Adj. R^2	0.335	0.260	0.243	0.167*	0.152	0.141
Collinearity	8.480	5.280	16.825	3.44*	7.881	3.479
p (Jarque–Bera)	0.340	0.193	0.612	0.103*	0.261	< 0.001#
p (Koenker–Bassett)	0.111	0.048#	0.753	0.007*	0.759	0.130
p (Moran's I)	0.492	0.755	0.268	0.019*	0.940	0.564
Coefficients						
% Medium-intensity development (per 10%)	0.019	0.049		0.018		
Per capita Income (per $10000)	−0.015					
% Over age 65 (per 10%)		0.605	0.035		0.044	
% Asian (per 10%)			0.039			
% Under age 5 (per 10%)			0.097			
% White (per 10%)			−0.005			
% Pacific Islander (per 10%)					0.436	
% Below high school education (per 10%)						0.282
Other terms in model						
Spatial lag				0.297		
Constant	0.932	0.715	0.886	0.576	0.876	0.870

* The diagnostic information shown for Phoenix is for the regression model excluding the spatial lag term; the coefficients are shown for the final model that does include the spatial lag term.

The significant tests for residual normality in St. Louis and residual heteroskedasticity in Minneapolis are caused by one single outlier value that did not have a substantive impact on the regression.

Phoenix), higher percentages of residents of Pacific Islander heritage (in Seattle), and higher percentages of residents without a high school education (in St. Louis). The environmental variable associated with higher risk was percentage of medium-intensity development (in Boston, Minneapolis, and Phoenix). A spatial lag term for the dependent variable was included in the model for Phoenix only, as the original model without a spatial lag term showed significant spatial autocorrelation in the residuals. This effect was reduced but not completely removed by adding a spatial lag term to the model.

3.3. Spatial regression with principal components

Significant associations between principal components of the demographic and environmental factors and intra-city heat-related mortality rates were also evident in all cities (Table 4). On average, six principal components were extracted for each city that represented 81% of the variance of the original pool of 25 variables (see Supplemental materials for loadings tables). The principal components regression models explained 6.7–27.7% of the intra-city mortality patterns. In all six cities the first principal component was reflective of socioeconomic status, and this

Table 4
Summary information for regression models generated for each city using city-specific principal components of the suite of explanatory variables shown in Table 2. The complete loadings matrix for the principal components, each of which briefly characterized in this table, can be found in Supplemental materials.

	F	p	Adj. R^2	Terms included	Coefficient
Boston	7.311	< 0.01	0.243	PC1 (low education, high poverty, high-intensity development, nonwhite)	0.022
				PC2 (low-intensity development, children)	0.022
				PC4 (older housing, medium and low-intensity development)	0.021
Minneapolis	10.494	< 0.01	0.277	PC1 (old housing, low education, poverty)	0.041
				PC2 (elderly)	0.035
				PC3 (low-intensity development)	−0.034
				PC5 (natural land cover, other races)	0.024
Philadelphia	10.388	< 0.01	0.176	PC9 (Asian, Pacific Islander, elderly, older homes)	0.019
Phoenix	9.503	< 0.01	0.219	PC1 (low education, high poverty, high public assistance, high nonwhite)	0.030
				PC2 (high living alone, medium-intensity development)	0.032
				PC4 (high home values, high income, forest land cover)	−0.020
Seattle	5.227	0.03	0.067	PC1 (high public assistance, low education, high nonwhite, low income)	0.018
St. Louis	8.52	< 0.01	0.113	PC1 (high poverty, high public assistance, low education, low income)	0.029

component was included in regression models in five of the six locations.

Three principal components (PCs) were included in the model for Boston. A positive relationship was found for Boston PC2, which had strong positive loadings for percentage of children under age 5 and low-intensity development, and strong negative loadings for one person households and high-intensity development. Boston PC1 was also included in the model with a positive association. This component loaded positively on percent below poverty, percent without a high school education, and percent of several nonwhite races, and loaded negatively on open space, per capita income, and percent over age 65. The third and final term included in the model was Boston PC4, which had strongest positive loadings on medium-intensity development and old housing. This component was also positively related to areas with higher heat-related mortality. The model for Boston explained 24.3% of the heat-related mortality pattern.

Four PCs were included in the model for Minneapolis, which accounted for 27.7% of the spatial variability in heat-related mortality rates. Minneapolis PC1 had a positive regression coefficient and loaded positively on percent below high school education, with public assistance, below poverty, and medium-intensity development, and loaded negatively on percent white. Minneapolis PC2 loaded positively on percent elderly and percent living alone and negatively on percent under age 5; this component also had a positive regression coefficient. Minneapolis PC3 was inversely related to areas with high mortality. PC3 is low in places with high percentages of open space and low-intensity development. Minneapolis PC5 loaded strongly on percent from other races and water, wetland, and forest land cover types and was positively associated with high mortality zones.

No model was generated for Philadelphia using only PCs selected from the original search criteria as none of the partial significances were less than 0.05. However, when additional PCs were tested, Philadelphia PC9 (2.9% of original variance, $\lambda = 0.676$) was significantly positively related to the mortality pattern and explained 17.6% of the spatial variability. Loadings for this PC were strongest and positive for percent Asian, percent Pacific Islander, percent over age 65, and percent of homes built before 1970 or earlier.

Three PCs were included in the model for Phoenix and accounted for 21.9% of the spatial variability in heat-related mortality. Phoenix PC2 had high positive loadings for percent living alone, percent living alone over age 65, and medium-intensity development. PC2 had a positive regression coefficient. PC1, which had high positively loadings with percent below high school education, percent with public assistance, percent below poverty, and percent nonwhite, also had a positive coefficient. A negative relationship was found with Phoenix PC4, which had positive loadings for median housing value, per capita income, and forest land cover type.

The model for Seattle included only one variable, Seattle PC1. This component was strongly positive loaded on percent with public assistance, percent below high school education, and percent of nonwhite races, and negative loaded on per capita income and median housing value. PC1 had a positive regression coefficient and explained 6.7% of the spatial variability in heat-related mortality.

The model for St. Louis also included only one variable, St. Louis PC1. This model explained 11.3% of the spatial mortality pattern. PC1 was strongly positively loaded on percent below poverty, percent with public assistance, and percent below high school education. It was negatively loaded on median housing value, per capita income, and percent white.

Table 5
Summary information for regression models generated for data combined from all cities, analogous to the city-specific models shown in Tables 3 and 4.

	All cities (orig.)	All cities (PCs)
Model summary		
N	417	417
F	29.371	30.738
p	< 0.01	< 0.01
Adj. R^2	0.254	0.263
Coefficients		
% Medium-intensity development (per 10%)	0.014	
% Living alone over age 65 (per 10%)	0.033	
% Homes built before 1970 (per 10%)	0.517	
% White (per 10%)	−0.690	
% Pacific Islander (per 10%)	0.531	
PC1 (low education, high poverty)		0.039
PC2 (low children, high living alone)		0.016
PC4 (low-intensity development)		0.023
PC5 (high elderly, high grass and crops)		0.011
PC7 (high home values, high incomes, high Amer. Indian, high forest)		−0.013
Constant	0.869	0.919

3.4. All-city models

All postal codes from all cities were combined for the final set of regression models (Table 5). Using all subsets regression with the suite of original demographic and environmental variables, the optimal multiple regression model included five variables and explained 25.4% of the variance. Terms with a positive coefficient included percent of homes built prior to 1970, percent of elderly living alone, percent Pacific Islander, and percent medium-intensity development. Percent white was included in the model with a negative coefficient. The strongest cross-correlation between these independent variables was between percent white and percent of homes built before 1970 (−0.426); most of the correlations had absolute values less than 0.25.

Analogously to the individual city models, we extracted principal components from the all-city data set. Seven of the components (AllCity PCs) had eigenvalues greater than one and these PCs accounted for 75.4% of the variance in the original data set. Five AllCity PCs were included in a multiple regression model that explained 26.3% of the variance in heat-related mortality across the entire data set. Four terms in the model, AllCity PC1, PC2, PC4, and PC5 had positive regression coefficients. AllCity PC1 had strong positive loadings on percent below high school education, percent below poverty, and percent with public assistance. AllCity PC2 was loaded most strongly on percent under age 5 (negative) and percent living alone (positive). AllCity PC4 had a high positive loading for low-intensity development. PC5 had strong negative loadings on percent elderly and percent grass and crops land cover type. The fifth term included in the model, AllCity PC7, had a negative regression coefficient. This component had highest positive loadings for median housing value, per capita income, and percent American Indian, and a strong negative loading for percent forest land cover type.

4. Discussion

In cities where high summer temperatures lead to elevated mortality rates, there is significant spatial variability in sensitivity to heat. Mortality records from recent time periods spanning 14–26 years show that residents of certain portions of cities have been at greater risk of dying when extreme heat occurs. To the best of our knowledge, this study is one of the first to document such intra-city variability in risk using long-term health outcome records.

4.1. Inter-city variability

The modeled relationship between temperature and mortality (Fig. 1) for the cities evaluated is similar to that reported elsewhere—a U- or J-shaped curve where mortality rates increase at the lowest and highest temperatures (Curriero et al., 2002; Davis et al., 2003). Unsurprisingly, we found that the threshold temperature, defined as the lowest temperature at which mortality is significant different than observed for normal summer conditions, varies geographically. The highest threshold temperature was found for the warmest study location, Phoenix, while the lowest thresholds were in the coolest locations, Seattle and Boston. No threshold temperature was found for Atlanta, Georgia, a city with high temperatures and humidity that persists throughout much of the summer. Although the modeled relationship for Atlanta (not shown) was similar in shape to the other cities, the confidence interval for the estimated effect widened considerably at the highest temperatures. No relationship (or a weak one) between summer temperature and mortality has previously been reported for Atlanta and other locations in the southeastern United States (e.g., Curriero et al., 2002). Geographical variability in the threshold temperature is consistent with previous research, and it is believed that this variability arises because people in different locations physically and technologically adapt to their climate (e.g., Davis et al., 2003).

The threshold temperatures that we identified are lower than temperatures commonly used in heat-warning systems at which various public health intervention strategies are activated. In some of the cities we examined, the threshold temperature was only two degrees (C) above the summer mean temperature, which resulted in a large sample of days included as “hot.” This is an important contrast to draw between this research and others that use more stringent criteria to identify extreme heat days such as the 95th or 99th percentile summer temperature (e.g., Gosling et al., 2007; Anderson and Bell, 2009). From a statistical standpoint, temperatures only a few degrees above normal summer conditions are associated with elevated mortality rates and should be considered when evaluating the total health burden related to high temperatures and when projecting future health impacts under climate change. However, we do not advocate that these lower thresholds be used in public health alert systems, as alarming the public too frequently may result in diminished alert effectiveness.

4.2. Intra-city variability

There is strong evidence in support of the hypothesis that heat-related mortality is spatially variable within urban areas. In each of the six cities examined, significant increases in mortality when temperature exceeded the city-specific threshold were confined to only a portion of the postal codes comprising each study area. This study joins a small but growing body of research documenting such inter-city variability based on historical medical data (e.g., Schuman, 1972; Smargiassi et al., 2009; Vaneckova et al., 2010; Laaidi et al., 2012; Reid et al., 2012; Johnson et al., 2012; Harlan et al., 2013). A portion of this previous work has focused on single locations and/or single heat events, whereas here we have included long records of mortality data for multiple locations. Within the cities we investigated, the postal codes associated with high mortality rates on hot days tend to be consistent from one year to the next (Hondula and Davis, 2014), but continued examination of temporal factors is important in light of a growing body of literature documenting changes in heat-related sensitivity over time (e.g., Davis et al., 2003; Petkova et al., 2014).

4.2.1. Spatial regression with original variables

We found a significant relationship between the spatial pattern in mortality and various potential explanatory variables in each city, but the specific variables included in the optimal multiple regression model varied from one city to another. Only two variables were included in the model for more than one city (percent land cover with medium-intensity development and percent of elderly residents). Thus, although certain variables may be important determinants of heat-related risk regardless of location, at this spatial scale, the strongest associations between socioeconomic and environmental factors and mortality outcomes are inconsistent from place to place. Because patterns in socioeconomic and environmental factors vary between cities, there may be interactions and/or competing effects in one city that are not present elsewhere.

The key risk factors for Boston were per capita income and percent medium-intensity development. Six contiguous postal codes were associated with high heat-related mortality rates in the southern part of the study area that include areas in and near Roslindale and Mattapan. These postal codes had both high percentages of developed land and low per capita incomes. There are sharp contrasts between these areas in terms of percentage of high-intensity land cover type, indicating that high heat-related mortality rates are possible even in places that are not characteristic of a central business district. Moving northward into the city center, two adjacent postal codes between Roxbury and the city center were associated with significant increases. Both areas had high percentages of medium- and high-intensity developed land, but they were markedly different in terms of per capita income (more than \$60,000 in the more eastern of the two versus below \$22,000 in the more western). In the higher-income area, approximately 15% of residents live below the poverty line, and nearly 60% of households have only one person, so it is possible that these variables are contributing to higher risk in this area despite higher overall wealth. Per capita income in the East Boston area in the postal code containing Logan airport is among the lowest in the entire city. In the remaining postal codes with high mortality rates to the north and west of the city center, incomes are typically slightly below the citywide mean and percentage developed land is higher than the citywide mean, although there are other areas with similar characteristics without high-risk. Model-predicted mortality in these locations was below observations. Thus there may be a separate risk factor here that is not captured in the multivariate regression.

Statistically significant elevations in mortality on above-threshold days are evident in areas scattered throughout the Minneapolis region, but the regression model indicates two predominant covariates: percent medium-intensity development and percent elderly. Postal codes that feature the characteristics indicated by the model and high heat-related mortality include those immediately south of the central business district of Minneapolis, the postal code to the south and east of Edina, and the area between Minneapolis and Columbia Heights to the north. The other postal codes with significant elevations in mortality on above-threshold days do not fit the model well, as they have both percent elderly and percent medium-intensity land cover at or below the regional mean. Two areas that particularly poorly fit include the postal code southeast of Andover in Anoka and Ramsey counties, and the postal code including the lake district in the far west of the study region. It is difficult to build a hypothesis for sources of high-risk in these regions based on the literature and data examined in this study. With respect to the westernmost postal code, it is possible that recreation on hot days is a driver of elevated risk, as one study from Toronto showed higher ambulance call-outs near the lakeshore where people sought relief from the heat (Bassil et al., 2009).

Nearly half of the postal codes in Philadelphia County are associated with statistically high mortality rates on above-threshold days, and the density of high-risk locations is greatest in the southern and central portions of the County. Four factors were found to be associated with mortality risk, including percent elderly, percent Asian, percent under age five, and percent white (protective). Postal codes with high percentages of elderly residents and higher heat-related mortality rates are located on the perimeter of the study region, including Northwest Philadelphia and residential portions of South Philadelphia. The two postal codes with the highest percentage of elderly residents, however, located in the most northern part of the County, were not associated with elevated risk. The five postal codes with the highest percentages of Asian residents were all associated with statistically high heat-related mortality, three of which are located in the southeasternmost portion of the County. The postal codes in the center of the county associated with high mortality had highest rates of children under age 5 and lower percentages of white residents. These postal codes also have the lowest per capita incomes in the County, but income alone was not found to be a significant predictor of mortality. The westernmost postal code, north of Upper Darby Township, and the postal code containing Southwest Philadelphia did not fit the model well.

The variable that explained the greatest amount of variability in heat-related mortality rates across Phoenix was percent medium-intensity land cover; the addition of other variables did not significantly improve the model and led to higher BIC. Of the 19 postal codes with 50% or more medium-intensity land cover, four are associated with high heat-related mortality rates located in and near Glendale. This contrasts with only one of the 33 postal codes with 20% or less medium-intensity land cover having high-risk. The highly developed, high-risk areas are also associated with low income, and previously a high correlation has been found for Phoenix between development/urban heat island effects and poverty status (Harlan et al., 2007). Three of the remaining high-risk postal codes located near Paradise Valley, east Gilbert, and Sun City West, are associated with high percentages of elderly residents. The high-risk area between Mesa and Phoenix has among the lowest mean per capita incomes in the region. Three other high-risk zones located in and near Scottsdale do not fit the model well, nor do they seem to be associated with high-risk related to any other variables we considered. Higher counts of heat-related deaths in the Scottsdale area were also observed by Harlan et al. (2013). It is important to acknowledge that the model for Phoenix was the only one that also included a term to account for spatial autocorrelation in the dependent variable, and these spatial effects were not completely removed with the addition of a spatial lag term, leaving the possibility that the model is not properly specified.

Postal codes in Seattle with higher percentages of elderly residents and Pacific Islander residents were associated with higher risk. High-risk postal codes in the central and western portions of the study region all are associated with some of the highest rates of Pacific Islander residents in the Seattle area. However, it should be noted that Pacific Islanders represent no more than 2.5% of the population in any single postal code. The areas near Bellevue, east of Seattle, and Fort Lawton to the northwest have higher percentages of elderly residents and higher heat-related mortality rates. High-risk areas near the central business district and to the south in eastern Auburn have lower average per capita incomes and more medium- and high-intensity development, which could be contributing to higher risk, but there is little or no association between risk and either of these variables across the entire Seattle area. The high-risk area east of Bellevue has a high percentage of Asian residents, which was also found for certain high-risk areas in Philadelphia.

Two clusters of high-risk zones are evident in the St. Louis area, one including postal codes in and immediately proximate to the city proper on the eastern edge of the study region, and another in outlying areas west of Chesterfield. The only explanatory variable included in the model was percent of residents with less than a high school education, which explained over 14% of the spatial variability. Three of the high-risk postal codes located near the city proper are among the ten lowest in terms of educational attainment. Other high-risk postal codes in this area include high rates of poverty (25% or more), and high prevalence of homes constructed before 1940 (60% or more). Throughout the city proper, there are many postal codes with low educational attainment that are not associated with higher risk, and future work might help understand those factors that account for this differentiation. On the western edge of the study region, the postal code just west of Chesterfield has the highest level of educational attainment in the region and is thus an outlier for the model, causing the normality in error diagnostics to return a significant value (Table 3). This is a highly affluent area with few elderly and newer homes, and thus the source of risk is unclear. The postal code immediately to the south has lower educational attainment and income, although both values are close to the regional mean.

Across all cities, the associations between the model-selected variables and spatial mortality patterns are generally consistent with expectations one would derive from the literature. Environmental factors were included in models in three of the six study cities. Mortality is higher in places with greater percentages of developed land in Boston, Minneapolis, and Phoenix, which aligns with other studies investigating Montreal, Paris, Phoenix, and Chicago (Smargiassi et al., 2009; Laaidi et al., 2012; Johnson et al., 2012; Harlan et al., 2013). The associations we found for built environment characteristics were for medium-intensity development; high-intensity development was not included in the model for any city. For some cities this variable was intentionally excluded because of non-normality (many postal codes have zero high-intensity development) and/or high collinearity with other variables that had stronger associations with the mortality pattern. High-intensity development may not be a strong covariate with heat-related mortality because many of the most highly developed areas within cities have commercial districts with few residents or sometimes feature very expensive residences. The growing availability of high-resolution intra-urban environmental monitoring networks (e.g., Overeem et al., 2013; Chapman et al., 2014) will enhance our ability to investigate the contribution of urban microclimate variability to human health outcomes associated with extreme heat exposure. The extent to which these data enhance modeling approaches like the one we have employed here is unclear, however, as Guo et al. (2013) demonstrated that a time series approach for heat-related mortality without information about intra-urban temperature variability performed equally as well in predicting heat-related health outcomes in Brisbane, Australia. More urgently, new research is required to understand how intercity differences in the built environment are associated with differences in thermal exposure for residents (e.g., Kuras et al., 2015) but the results of this study support the hypothesis that built environment factors contribute to spatial variations in heat-related mortality in certain locales. Some evidence to the contrary in the literature was conducted at coarser spatial scales (Sheridan and Dolney, 2003; Hattis et al., 2012).

Demographic factors were included in models for all six study cities and were also generally consistent with previous research.

Income and age, variables that commonly appear as risk factors in the heat-related literature, also appeared as key predictors in this study. Per capita income was inversely associated with heat-related mortality rates in Boston. Income is believed to be an important determinant of risk related to heat because those living with low incomes and/or in poverty may not be able to afford air conditioning, which is among the best means of protecting oneself during periods of extreme heat (Semenza et al., 1996; Naughton et al., 2002). Air conditioning data at the postal code scale was not available for this study, and thus income may be serving as the most appropriate proxy measure. It is surprising that income or income-related measures were not included in the models for more cities. Postal codes with higher percentages of elderly residents or higher percentages of elderly residents living alone were associated with higher heat-related mortality in Minneapolis, Philadelphia, and Seattle. Postal codes with more young children also had higher mortality rates in Philadelphia. These findings are also consistent with previous research (e.g., McGeehin and Mirabelli, 2001; Basu and Ostro, 2008).

A number of other demographic factors were also included in the multiple regression models that have received less attention in the literature. Postal codes with a lower percentage of white residents were associated with higher heat-related mortality rates in Philadelphia, as were postal codes with higher percentages of Asian residents. In Seattle, postal codes with higher mortality rates had higher percentages of residents with Pacific Islander heritage. The effect of race on heat-related risk has been examined in some research that has drawn mixed conclusions (e.g., Kalkstein and Davis, 1989; O'Neill et al., 2003; Hattis et al., 2012). In Philadelphia, we found that the percentage of white residents was highly correlated with a number of other variables including income and educational attainment, and thus for this city the variable selected may be serving as a proxy for larger-scale demographic patterns (Hondula et al., 2012). Lower percentages of educational attainment were found to be associated with higher risk in St. Louis, which may be reflective of either the benefits of an educated public in understanding risk related to heat and the appropriate precautionary measures to take. We note that education and income were well correlated in St. Louis.

4.2.2. *Comparison to principal components regression*

Principal components regression yielded statistical models that explained, on average, less variance in each city's spatial pattern in heat-related mortality than those generated using the original explanatory variables. The variance explained from the principal components models was greater in Minneapolis (by 1.7%) and Phoenix (by 5.2%). In the other four cities the principal components models explained 2.8–9.2% less variance. Thus while principal components analysis offers a theoretical statistical advantage in that all of the explanatory variables are orthogonal, there was no strong evidence that these terms were any more closely associated with spatial variability in mortality than combinations of the original variables that were tested for collinearity.

The principal components included in the regression models do provide a different perspective on the postal code level characteristics associated with variability in heat sensitivity, as they more comprehensively capture demographic and environmental patterns that are not well-represented by any single variable. For example, Boston PC2, which had high positive loadings for low-intensity development and percent children under five and strong negative loadings for percent living alone and higher density development types, seems to be clearly capturing a contrast between single urban dwellers and suburban families. In the regression model the latter of these two patterns was associated with increasing risk of heat-related mortality. This association was not evident from the single variable analysis, which yielded medium-intensity development and low incomes as key predictors.

Minneapolis was one of the two cities where the principal components regression had a higher adjusted R^2 than the original multiple regression, but the key predictors were found to be largely the same. Areas with more elderly residents and areas with higher development are associated with higher risk. The PC regression did yield another pattern not evident from the original regression, in that places with old housing, uneducated residents, and residents living in poverty, were also at higher risk. Based on the Minnesota model, PC regression appears to be advantageous because it allows for different combinations of variables to be associated with variability in risk. A similar result was found for Phoenix, where the original model included only one variable but the PC regression yielded three distinct spatial patterns associated with higher risk.

A contrasting example, however, emerged in Philadelphia, where none of the principal components with eigenvalues greater than one were significantly related to the intra-city heat-related mortality pattern. Compared with the original multivariable model for Philadelphia, this PC outcome demonstrates that it is possible that none of the predominant spatial patterns in demographic or environmental variables are associated with spatial variability in risk. Only in the more subtle spatial pattern represented by PC9 was an association with heat-related risk found. Interestingly, the variables represented by Philadelphia PC9 (percent Asian, percent elderly, and old housing) share many commonalities with the variables included in the original regression (percent elderly, percent Asian, percent under age 5, and percent nonwhite).

As was true for the original multivariable models, the portion of variance explained by principal components in Seattle and St. Louis was the lowest of the six cities examined. The key factors associated with higher risk were found to be different in both cases. In Seattle, the PC most closely associated with the spatial variability in risk was representative of general deprivation and prevalence of minority races, whereas the original model included percent elderly and percent of one specific race. In St. Louis, education was the key factor from the original models, but was not highly loaded on the component included in the PC regression. Instead, the component captured a pattern related to collocated poverty, old houses, and prevalence of minority races. As both the original and PC models accounted for relatively low percentages of the variance, it is unsurprising that the set of key predictive factors differs, as there appeared to be no predominant spatial pattern to capture regardless of collinearity among the independent variables.

The principal components also provide additional perspective on the results of the regression using the original set of variables. The procedure included tests for collinearity among independent variables included in the models, but there was no consideration for collinearity of excluded variables. Thus, the possibility exists that variables included in the model are highly collinear with excluded variables and may be representative of a different pattern than might be suspected simply from examining only the included variables.

Originally included in the model for Boston were per capita income and medium-intensity development. Both of these variables have high loadings on Boston PC1, which is representative of general socioeconomic status and racial variability. Additional variables with high loadings on Boston PC1 include educational attainment, public assistance, poverty, and race. All of these factors could contribute to areas with lower incomes and more intense development being associated with higher risk. Boston PC1 had a strong negative loading for percent elderly, providing an example of a relationship where places with fewer elderly have higher risk because of other factors. Per capita income also had a high loading on Boston PC2, representative of the suburban/urban contrast previously discussed. As the set of variables with high loadings on PC2 substantively differs from those with high loadings on PC1, there are at least two separate income-related patterns associated with higher risk in Boston.

Variables included in the original model for Minneapolis were percent over age 65 and medium-intensity development. Percent elderly had a high loading on Minneapolis PC2, which was the predominant variable contributing to that component. Minneapolis PC2 was also included in the principal components regression, and accordingly spatial variability in the prevalence of elderly is a pattern related to risk largely independent from other variables. Conversely, percent medium-intensity development had a high loading on Minneapolis PC1, as was the case for median housing value, per capita income, educational attainment, public assistance, poverty, and race. Medium-intensity development may be included in the Minneapolis model as an indicator of overall socioeconomic status. Minneapolis PC5 most heavily loaded on land cover types and percent of other races, none of which were included in the original models. The positive association between PC5 and mortality is counterintuitive when considering the land cover type loadings alone—this highlights a limitation of the principal components approach and ecological design in general with limited understanding of any underlying mechanisms that cause land cover type and racial distribution to be spatial correlated in this location.

Of the original variables included for Philadelphia, three (percent elderly, percent children, and percent white) have high loadings on Philadelphia PC1, which like Minneapolis is an indicator of overall socioeconomic status. The loading for elderly on this component is negative, however—there are few elderly in the most economically disadvantaged areas of Philadelphia. But risk is high in these places, captured by in the inclusion of percent under five years old in the original regression, as this variable is highly correlated with many other socioeconomic indicators. Philadelphia PC1 may not have been included in the principal components regression because, with respect to heat-related mortality, it includes variables with a contradictory relationship to risk. The same could be said for Philadelphia PC4, which has high loadings on percent elderly and percent white, two variables with opposite signs in the original multivariate regression. Philadelphia was the only city where percent Asian was included as a predictor variable; percent Asian loads highly on Philadelphia PC2. Places with high scores for PC2 are highly urbanized (high-intensity development has a positive loading while medium- and low-intensity have a negative loading), higher median housing values, more people living alone, and few children. If being Asian is not a true driver of heat-related mortality risk, the inclusion of percent Asian in the regression model may be highlighting built environment effects or living alone as important factors.

The only variable included in the model for Phoenix was percent medium-intensity development, which is highly loaded on Phoenix PC1 and PC2. As is the case in other cities, Phoenix PC1 is representative of general socioeconomic status. The finding that risk is high in places with both more intense development and lower socioeconomic status is consistent with previous research (Harlan et al., 2007), but also hinders attributing risk to either factor independently in this location. A different spatial pattern, evident in Phoenix PC2, associates medium-intensity development, percent living alone, and percent elderly. There may be two separate spatial patterns at work that both associate medium-intensity development with heat-related risk even if built environment characteristics are not the underlying cause. The reverse is also possible, that those who are more socioeconomically disadvantaged and those living alone may not be at higher risk for those reasons, but are at greater risk because they live in more developed locations where their exposure to heat is likely to be more severe because of urban heat island effects.

Percent elderly and percent Pacific Islander were the two variables originally associated with risk variability in Seattle. As was true in Minneapolis, the highest loading for percent elderly is on a component (Seattle PC4) that is mostly reflective of spatial variability in percent elderly. Thus percent elderly is a unique spatial pattern in Seattle. On the other hand, percent Pacific Islander has a high loading on PC1, which, like other cities, is largely a socioeconomic indicator.

The original model for St. Louis included percent below high school education, which loads highly on St. Louis PC1. As is the case elsewhere, places with high PC1 scores in St. Louis have lower socioeconomic status and a greater prevalence of older dwellings. Percent medium and high-intensity development are also highly loaded on St. Louis PC1, indicating similarity to Phoenix that the most socioeconomically disadvantaged are living in places with more intense urbanization.

Common to five of the six cities examined was the inclusion of the first component in the principal component regression model. High scores for this component were associated with lower socioeconomic status and higher mortality rates on threshold-exceeding days. Thus, in general, lower socioeconomic status was associated with higher risk in a majority of the locations we examined. However, the specific variables contributing to this component varied from one city to another, which makes it difficult to construct an overall socioeconomic indicator appropriate for each city. For example, places with high scores on PC1 tended to have more homes built before 1940 or 1970 in most locations, but the relationship was weaker in Boston and hardly evident in Seattle. Isolation rates were also highly variable across PC1: there was a strong positive loading (0.65) for percent living alone on Minneapolis PC1, but a negative loading (-0.40) for percent living alone on Philadelphia PC1. It may be for this reason that Philadelphia PC1 was not included in the Philadelphia principal components regression model, whereas individual variables that covary with socioeconomics (percent under five, percent nonwhite) were included in the original models. Spatial variability in heat-related mortality is related to predominant spatial patterns in the independent variables across most of the cities examined, but an even more detailed understanding of interactive effects of the independent variables on heat-related risk could help generate different statistical models that account for greater portions of the spatial variability in mortality on hot days.

4.3. All-city models

The all-city models initially yielded a surprising result, in that the percentage of variance explained by both the original multivariate model and the principal components model was superior to that of many of the individual city models. This seemed to indicate universality in the risk factors of postal codes with high heat-related mortality, sharply contrasting with the finding from the city-specific models that the key risk factors varied from one place to another.

Subsequent analysis, however, revealed that the strength of the all-city models likely emerges from inter-city differences in the independent variables rather than true universal relationships between certain risk factors and heat-related mortality for individual postal codes. Analysis of variance revealed that model predicted values and residuals significantly varied by city from both the original multivariate regression and the principal components regression (see Supplemental materials). Thus, it seems that the all-city models are not properly specified, and the outcome of this analysis is analogous to an ecological fallacy, drawing inferences about variability between locations or individuals based on the results of one ecological study at one location (Piantadosi et al., 1988). The notion of an all-city model remains an interesting point for future exploration, but likely requires a different modeling approach, more contiguous spatial regions, and/or standardization of variables from one city to another. We cannot conclude from this analysis that there are informative universal predictors of places of high heat-related mortality across the cities we examined.

4.4. Synthesis

The key point arising from the results is that the specific predictor variables that best explain the spatial variability in heat-related risk vary from one city to another. This was evident by comparing models across cities using the original explanatory variables as well as those generated from city-specific principal components. This confirms the findings of a number of other researchers who have investigated intra-city (or inter-region) variability in heat-related risk, generally focusing on only one location. In Massachusetts, percent African–American and percent elderly population were found to be associated with variability in heat-related mortality rates across 29 districts comprising the state (Hattis et al., 2012). Johnson and Wilson (2009) found that poverty and urban heat island effects were key predictor variables for one heat event in Philadelphia. In a separate study of heat-related mortality in Philadelphia, home values and percentage of African–American residents were identified as the best predictors (Uejio et al., 2011). No relationship was found between socioeconomic status, vegetation cover, and variability in heat sensitivity among the elderly in Sydney, Australia (Vaneckova et al., 2010). Key risk variables reported in one study of Phoenix heat-related ambulance calls included intra-city temperature variability, percentage of African–American and Hispanic residents, isolation, and household vacancy (Uejio et al., 2011). This sample of studies represents the diverse methodologies, dependent variables, time periods, spatial scales of analysis, and conclusions reached regarding the variables most closely associated with spatial variability in heat-related risk. Here we show that different variables best account for the spatial pattern in mortality across different cities using consistent methods.

A related line of inquiry into spatial variability in extreme heat risk involves the development of heat vulnerability indices. In an evaluation of one such index across several different locations, Reid et al. (2012) found that the association between hospitalizations and the index varied from place to place and across disease types. Johnson et al. (2012) demonstrated that a heat vulnerability index developed specifically on spatial patterns in social and environmental data for Chicago explained a large portion of spatial variability in mortality arising from the Chicago 1995 heat wave. A similar association was found using a city-specific index developed for Phoenix focusing on deaths directly related to heat exposure over a nine-year time period (Harlan et al., 2013). The indices used in these two studies, and thus the role of key predictor variables, varied, as they were developed based on each location's demographic and environmental data. We have made no assumptions regarding the potential interactive or competing effects of variables that may have a similar spatial pattern, as is done when calculating a vulnerability index *a priori*, and we encourage more research aimed at understanding the best manner in which to design such indices and measure their value and reliability. It should be noted, however, that mismatched spatial patterns between heat vulnerability indices and heat–health outcomes do not necessarily limit the utility of either in developing adequate response measures. It may be the case, for example, that a neighborhood with high heat vulnerability but relatively low adverse heat–health outcomes has underlying resiliency to extreme heat that is not well-captured in our current understanding of vulnerability. Learning the adaptation mechanics at work in this neighborhood could be helpful for other communities; interest in this neighborhood might not have occurred without the simultaneous investigation of theoretical vulnerability and observed health outcomes.

There are several components of this study that could be addressed differently in future work. To start, this study examined all single days that occurred above a given temperature threshold. We did not treat periods of consecutive above-threshold days (“heat waves”) any differently. Some research has demonstrated some added effect on mortality when high temperatures occur several days in succession (e.g., Anderson and Bell, 2009, 2010), although this effect has generally be found to be small when compared to the overall effect of temperature (e.g., Hajat et al., 2006; Gasparrini and Armstrong, 2011). There was also no consideration of the date when high temperatures occurred, but it may be the case that intra-city variability in risk is different for early-season versus late-season heat events if there is geographical variability in acclimatization rates.

We did not examine modification of effects by air pollutants, which is a topic of continued active discussion in the literature (Theoharatos et al., 2010; Ren et al., 2011). If the spatial pattern in mortality risk on hot days varies between periods of good and poor air quality, health officials may be able to more efficiently target their intervention strategies. Future work may also explore how different temperature thresholds, spatial units, exposure variables, and lagged and displacement effects are manifest with respect to spatial variability in heat-related mortality. The total impact of heat on all-cause mortality may be underrepresented in this manuscript as we only examined the effect on the day on which above-threshold temperatures were observed. It would also be valuable to compare intra-city response in a range of health events beyond excess all-cause mortality (e.g., specific causes of death, hospitalizations, emergency department visits, and/or ambulance dispatches). The methodological framework we have presented could serve as a starting point for such research and the extension of this type of investigation to other locations.

Other modeling techniques like geographically weighted regression or regression trees could reveal in more detail variability in the importance of certain variables across space within cities. The fact that, in some cities, there were contrasts in the key factors that emerged from the original multivariate regression and those from principal components regression, suggests that there are likely complex interactions and spatial variability in the importance of certain predictors of risk within each city. Different modeling approaches like those suggested above may be able to better capture some of these patterns and help identify different combinations of factors associated with high risk. There may also be absolute and relative thresholds in the relationships between certain factors and risk of heat-related mortality, warranting a modeling approach that can account for nonlinear relationships. The ecological design of the study limits the applicability of results to the spatial framework presented herein and should not be extended to other spatial scales or to consideration of individuals (Piantadosi et al., 1988). Conclusions based on characteristics of places versus the specific pathways by which individuals suffer adverse heat–health outcomes could lead to misrepresentation of the most effective intervention strategies at the personal level (Duneier, 2006; Kuras et al., 2015). Public health officials in the jurisdictions we examined might use the information provided herein as a starting point for the development of localized intervention strategies. Ideally, these strategies would be enhanced by contextual information drawn from their own expertise of the neighborhoods, social dynamics, and resource accessibility in the populations they serve that are not possible to understand based on retrospective analysis of administrative data alone.

5. Conclusions

During the study period, days with temperatures above a city-specific threshold have been associated with statistically significant increases in mortality rates in six major U.S. metropolitan areas. On these hot days, all-cause mortality rates increase by several percent in each city, with higher increases associated with higher temperatures. However, the mortality rate on hot days within each city is variable from one place to another, and statistically significant increases in mortality are confined to less than 50% of the postal codes comprising each municipality we investigated. Spatially targeted short- and long-term intervention measures may be effective in reducing the public health burden related to extreme heat. Furthermore, demographic and environmental variables are associated with spatial variability in risk, enabling public health officials and city planners to accurately target certain populations and design strategies with heat-related mitigation measures. The specific variables most closely associated with spatial variability in heat-related risk vary from one city to another. Postal codes with more elderly residents, lower per capita income, more Pacific Islanders, more Asians, fewer whites, more children, and more developed area were found to be associated with higher risk, although no more than four of these factors were found to be the most closely associated variables in any one city. Principal components regression yielded an overlapping but non-identical set of predictor variables, but these models generally accounted for lower portions of the intra-city variance.

The use of daily georeferenced health data offers the opportunity to directly identify places within large metropolitan areas where the risk has historically been greatest on hot days. The areas within cities that appear to merit targeted intervention measures may differ from those one would identify using traditional notions of the drivers of spatial variability in heat-related risk that have not been verified by the examination of fine-scale mortality records within urban areas.

Role of funding sources

Research support was provided by the National Science Foundation Graduate Research Fellowship Program, the U.S. Environmental Protection Agency Science to Achieve Results Graduate Research Fellowship Program, the Virginia Space Grant Consortium Graduate Research Fellowship, the Association of American Geographers Dissertation Research Grant program, the Department of Environmental Sciences at the University of Virginia, and the Virginia C. Piper Health Policy Informatics Initiative at Arizona State University. The funding sources had no role in collecting, analyzing, or interpreting data, writing of the report, and in the decision to submit the article for publication. The authors declare no competing financial interests.

Acknowledgements

The authors are grateful from research support provided by the entities listed above, and for valuable input from Adrian Barnett, Joacim Rocklöv, Larry Kalkstein, Evan Kuras, Karen Smoyer-Tomic, Matthew Trowbridge, John Porter, and Bruce Hayden. Additional thanks to the departments of health and offices of vital statistics in the states (AZ, GA, MA, MN, MO, PA, WA) that contributed data for this study. These agencies bear no responsibility for the methods, results, and conclusions presented in this manuscript.

Appendix A. Supplementary Information

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.envres.2015.02.033.

References

Anderson, B.G., Bell, M.L., 2009. Weather-related mortality: how heat, cold, and heat waves affect mortality in the United States. Epidemiology 20 (2), 205.
Anderson, G., Bell, M.L., 2010. Heat waves in the United States: mortality risk during heat waves and effect modification by heat wave characteristics in 43 U. S. communities. Environ. Health Perspect. 119 (2), 210–218.
Anselin, L., Syabri, I., Kho, Y., 2006. GeoDa: an introduction to spatial data analysis. Geogr. Anal. 38 (1), 5–22.
Barnett, A.G., Tong, S., Clements, A.C.A., 2010. What measure of temperature is the best predictor of mortality? Environ. Res. 110 (6), 604–611.
Bassil, K.L., Cole, D.C., Moineddin, R., Craig, A.M., Wendy Lou, W.Y., Schwartz, B., Rea, E., 2009. Temporal and spatial variation of heat-related illness using 911 medical dispatch data. Environ. Res. 109 (5), 600–606.
Basu, R., Ostro, B.D., 2008. A multicounty analysis identifying the populations vulnerable to mortality associated with high ambient temperature in California. Am. J. Epidemiol. 168 (6), 632–637.
Centers for Disease Control and Prevention (CDC), 2012. CDC's Building Resilience Against Climate Effects (BRACE) Framework. Available: ⟨http://www.cdc.gov/climateandhealth/brace.htm⟩ (accessed 01.09.14).
Chapman, L., Muller, C.L., Young, D.T., Warren, E.L., Grimmond, C.S.B., Cai, X.M., Ferranti, E.J., 2014. The Birmingham Urban Climate Laboratory: an open meteorological testbed and challenges of the smart city. Bull. Am. Meteorol. Soc.
Chow, W.T., Chuang, W.C., Gober, P., 2012. Vulnerability to extreme heat in metropolitan Phoenix: spatial, temporal, and demographic dimensions. Prof. Geogr. 64 (2), 286–302.
Curriero, F.C., Heiner, K.S., Samet, J.M., Zeger, S.L., Strug, L., Patz, J.A., 2002. Temperature and mortality in 11 cities of the eastern United States. Am. J. Epidemiol. 155 (1), 80–87.
Davis, R.E., Knappenberger, P.C., Michaels, P.J., Novicoff, W.M., 2003. Changing heat-related mortality in the United States. Environ. Health Perspect. 111 (14), 1712.
Duneier, M., 2006. Ethnography, the ecological fallacy, and the 1995 Chicago heat wave. Am. Sociol. Rev. 71 (4), 679–688.
Ebi, K.L., Schmier, J.K., 2005. A stitch in time: improving public health early warning systems for extreme weather events. Epidemiol. Rev. 27 (1), 115–121.
Fry, J., Xian, G., Jin, S., Dewitz, J., Homer, C., Yang, L., Barnes, C., Herold, N., Wickham, J., 2011. Completion of the 2006 National Land Cover Database for the Conterminous United States. PE&RS 77 (9), 858–864.
Gasparrini, A., Armstrong, B., 2011. The impact of heat waves on mortality. Epidemiology 22 (1), 68.
Gosling, S.N., Lowe, J.A., McGregor, G.R., Pelling, M., Malamud, B.D., 2009. Associations between elevated atmospheric temperature and human mortality: a critical review of the literature. Climat. Chang. 92 (3–4), 299–341.
Gosling, S.N., McGregor, G.R., Páldy, A., 2007. Climate change and heat-related mortality in six cities Part 1: model construction and validation. Int. J. Biometeorol. 51 (6), 525–540.
Green, H., Gilbert, J., James, R., Byard, R.W., 2001. An analysis of factors contributing to a series of deaths caused by exposure to high environmental temperatures. Am. J. Forensic Med. Pathol. 22 (2), 196–199.
Guo, Y., Barnett, A.G., Tong, S., 2013. Spatiotemporal model or time series model for assessing city-wide temperature effects on mortality? Environ. Res. 120, 55–62.
Hajat, S., Armstrong, B., Baccini, M., Biggeri, A., Bisanti, L., Russo, A., et al., 2006. Impact of high temperatures on mortality: is there an added heat wave effect? Epidemiology 17 (6), 632–638.
Harlan, S.L., Brazel, A.J., Jenerette, G.D., Jones, N.S., Larsen, L., Prashad, L., Stefanov, W.L., 2007. In the shade of affluence: the inequitable distribution of the urban heat island. Res. Soc. Probl. Public Policy 15, 173–202.
Harlan, S.L., Declet-Barreto, J.H., Stefanov, W.L., Petitti, D.B., 2013. Neighborhood effects on heat deaths: social and environmental predictors of vulnerability in Maricopa County, Arizona. Environ. Health Perspect. 121 (2), 197.
Hattis, D., Ogneva-Himmelberger, Y., Ratick, S., 2012. The spatial variability of heat-related mortality in Massachusetts. Appl. Geogr. 33, 45–52.
Hondula, D.M., Barnett, A.G., 2014. Heat-related morbidity in Brisbane, Australia: spatial variation and area-level predictors. Environ. Health Perspect. 122 (8), 831–836.
Hondula, D.M., Davis, R.E., Leisten, M.J., Saha, M.V., Veazey, L.M., Wegner, C.R., 2012. Fine-scale spatial variability of heat-related mortality in Philadelphia County, USA, from 1983–2008: a case-series analysis. Environ. Health 11 (1), 1–11.
Hondula, D.M., Davis, R.E., Rocklöv, J., Saha, M.V., 2013a. A time series approach for evaluating intra-city heat-related mortality. J. Epidemiol. Community Health. http://dx.doi.org/10.1136/jech-2012-202157.
Hondula, D.M., Vanos, J.K., Gosling, S.N., 2013b. The SSC: a decade of climate–health research and future directions. Int. J. Biometeorol 58 (2), 109–120.
Hondula, D.M., Davis, R.E., 2014. The predictability of high-risk zones for heat-related mortality in seven U.S. cities. Natural Hazards. http://dx.doi.org/10.1007/s11069-014-1213-5.

Johnson, D.P., Wilson, J.S., 2009. The socio-spatial dynamics of extreme urban heat events: the case of heat-related deaths in Philadelphia. Appl. Geogr. 29 (3), 419–434.

Johnson, D.P., Stanforth, A., Lulla, V., Luber, G., 2012. Developing an applied extreme heat vulnerability index utilizing socioeconomic and environmental data. Appl. Geogr. 35 (1), 23–31.

Kalkstein, A.J., Sheridan, S.C., 2007. The social impacts of the heat–health watch/warning system in Phoenix, Arizona: assessing the perceived risk and response of the public. Int. J. Biometeorol. 52 (1), 43–55.

Kalkstein, L.S., Davis, R.E., 1989. Weather and human mortality: an evaluation of demographic and interregional responses in the United States. Ann. Assoc. Am. Geogr. 79 (1), 44–64.

Kilbourne, E.M., Choi, K., Jones, T.S., Thacker, S.B., 1982. Risk factors for heatstroke. Jama 247 (24), 3332–3334.

Koppe, C., Kovats, S., Jendritzky, G., Menne, B., Breuer, D.J., 2004. Heat Waves: Risks and Responses. Regional Office for Europe. World Health Organization.

Kuras, E.R., Hondula, D.M., Brown-Saracino, J., 2015. Heterogeneity in individually experienced temperatures (IETs) within an urban neighborhood: insights from a new approach to measuring heat exposure. Int. J. Biometeorol., 1–10. http://dx.doi.org/10.1007/s00484-014-0946-x.

Laaidi, K., Zeghnoun, A., Dousset, B., Bretin, P., Vandentorren, S., Giraudet, E., Beaudeau, P., 2012. The impact of heat islands on mortality in Paris during the August 2003 heat wave. Environ. Health Perspect. 120 (2), 254.

Loughnan, M.E., Tapper, N.J., Phan, T., Lynch, K., McInnes, J.A., 2013. A Spatial Vulnerability Analysis Of Urban Populations During Extreme Heat Events in Australian Capital Cities, Gold Coast, Australia: National Climate Change Adaptation Research Facility.

Lowe, D., Ebi, K.L., Forsberg, B., 2011. Heatwave early warning systems and adaptation advice to reduce human health consequences of heatwaves. Int. J. Environ. Res. Public Health 8 (12), 4623–4648.

McGeehin, M.A., Mirabelli, M., 2001. The potential impacts of climate variability and change on temperature-related morbidity and mortality in the United States. Environ. Health Perspect. 109 (Suppl 2), 185.

Michigan Public Health Institute (MPHI), 2011. Mapping heat vulnerability in Michigan. Available: ⟨https://www.michigan.gov/documents/mdch/Heat_Mapping_FY11_Final_Report_9.30.11_433139_7.pdf⟩ (accessed 01.09.14).

Minnesota Population Center, 2011. National Historical Geographic Information System: Version 2.0. University of Minnesota, Minneapolis, MN.

National Center for Health Statistics (NCHS), 2014. Deaths Attributed to Heat, Cold, and Other Weather Events in the United States, 2006–2010. Available: ⟨http://www.cdc.gov/nchs/data/nhsr/nhsr076.pdf⟩ (accessed 03.08.14).

Naughton, M.P., Henderson, A., Mirabelli, M.C., Kaiser, R., Wilhelm, J.L., Kieszak, S. M., McGeehin, M.A., 2002. Heat-related mortality during a 1999 heat wave in Chicago. Am. J. Prev. Med. 22 (4), 221–227.

O'Neill, M.S., Zanobetti, A., Schwartz, J., 2003. Modifiers of the temperature and mortality association in seven US cities. Am. J. Epidemiol. 157 (12), 1074–1082.

Overeem, A., Robinson, R., Leijnse, J.C., Steeneveld, H., P Horn, G.J., Uijlenhoet, B.K., 2013. Crowdsourcing urban air temperatures from smartphone battery temperatures. Geophys. Res. Lett. 40 (15), 4081–4085.

Pascal, M., Laaidi, K., Ledrans, M., Baffert, E., Caserio-Schönemann, C., Le Tertre, A., Empereur-Bissonnet, P., 2006. France's heat health watch warning system. Int. J. Biometeorol. 50 (3), 144–153.

Petkova, E.P., Gasparrini, A., Kinney, P.L., 2014. Heat and mortality in New York City since the beginning of the 20th century. Epidemiology 25 (4), 554–560.

Piantadosi, S., Byar, D.P., Green, S.B., 1988. The ecological fallacy. Am. J. Epidemiol. 127 (5), 893–904.

Reid, C.E., O'Neill, M.S., Gronlund, C.J., Brines, S.J., Brown, D.G., Diez-Roux, A.V., Schwartz, J., 2009. Mapping community determinants of heat vulnerability. Environ. Health Perspect. 117 (11), 1730.

Reid, C.E., Mann, J.K., Alfasso, R., English, P.B., King, G.C., Lincoln, R.A., Balmes, J.R., 2012. Evaluation of a heat vulnerability index on abnormally hot days: an environmental public health tracking study. Environ. Health Perspect. 120 (5), 715–720.

Ren, C., O'Neill, M.S., Park, S.K., Sparrow, D., Vokonas, P., Schwartz, J., 2011. Ambient temperature, air pollution, and heart rate variability in an aging population. Am. J. Epidemiol. 173 (9), 1013–1021.

RStudio, 2012. RStudio: Integrated development environment for R (Version 2.15.1). Boston, MA, USA.

San Francisco Department of Public Health, 2012. Heat Vulnerability Index by Census Block Group San Francisco, CA. Available: ⟨http://www.sfdph.org/dph/files/EHSdocs/Climate/HeatVulnerabilityIndex.jpg⟩ (accessed 01.09.14).

Schuman, S.H., 1972. Patterns of urban heat-wave deaths and implications for prevention: data from New York and St. Louis during July, 1966. Environ. Res. 5 (1), 59–75.

Schwarz, G., 1978. Estimating the dimension of a model. Ann. Stat. 6 (2), 461–464.

Semenza, J.C., Rubin, C.H., Falter, K.H., Selanikio, J.D., Flanders, W.D., Howe, H.L., Wilhelm, J.L., 1996. Heat-related deaths during the July 1995 heat wave in Chicago. N. Engl. J. Med. 335 (2), 84–90.

Sheridan, S.C., Dolney, T.J., 2003. Heat, mortality, and level of urbanization: measuring vulnerability across Ohio, USA. Clim. Res. 24 (3), 255–265.

Sheridan, S.C., Kalkstein, L.S., 2004. Progress in heat watch-warning system technology. Bull. Am. Meteorol. Soc. 85 (12), 1931–1942.

Smargiassi, A., Goldberg, M.S., Plante, C., Fournier, M., Baudouin, Y., Kosatsky, T., 2009. Variation of daily warm season mortality as a function of micro-urban heat islands. J. Epidemiol. Community Health 63 (8), 659–664.

Steadman, R.G., 1984. A universal scale of apparent temperature. J. Clim. Appl. Meteorol. 23 (12), 1674–1687.

Theoharatos, G., Pantavou, K., Mavrakis, A., Spanou, A., Katavoutas, G., Efstathiou, P., Asimakopoulos, D., 2010. Heat waves observed in 2007 in Athens, Greece: synoptic conditions, bioclimatological assessment, air quality levels and health effects. Environ. Res. 110 (2), 152–161.

Uejio, C.K., Wilhelmi, O.V., Golden, J.S., Mills, D.M., Gulino, S.P., Samenow, J.P., 2011. Intra-urban societal vulnerability to extreme heat: the role of heat exposure and the built environment, socioeconomics, and neighborhood stability. Health Place 17 (2), 498–507.

Vaneckova, P., Beggs, P.J., Jacobson, C.R., 2010. Spatial analysis of heat-related mortality among the elderly between 1993 and 2004 in Sydney, Australia. Soc. Sci. Med. 70 (2), 293–304.

Vaneckova, P., Neville, G., Tippett, V., Aitken, P., FitzGerald, G., Tong, S., 2011. Do biometeorological indices improve modeling outcomes of heat-related mortality? J. Appl. Meteorol. Climatol. 50 (6), 1165–1176.

Wolf, T., McGregor, G., 2013. The development of a heat wave vulnerability index for London, United Kingdom. Weather Clim. Extrem. 1, 59–68.

Wood, S.N., 2006. Generalized Additive Models: an Introduction with R, Vol. 66. Chapman & Hall, Boca Raton, FL.

Yardley, J., Sigal, R.J., Kenny, G.P., 2011. Heat health planning: the importance of social and community factors. Glob. Environ. Chang. 21 (2), 670–679.