BIOMETRICS 000, 000–000 DOI: 000
000 0000

A Bayesian Model Averaging Approach for Estimating the Relative Risk of Mortality Associated with Heat Waves in 105 U.S. Cities

Jennifer F. Bobb[1,*], Francesca Dominici[2], and Roger D. Peng[1]

[1]Department of Biostatistics, Johns Hopkins School of Public Health, Baltimore, MD 21205

[2]Department of Biostatistics, Harvard School of Public Health, Boston, MA 02115

*email: jfeder@jhsph.edu

SUMMARY: Estimating the risks heat waves pose to human health is a critical part of assessing the future impact of climate change. In this paper we propose a flexible class of time series models to estimate the relative risk of mortality associated with heat waves and conduct Bayesian model averaging (BMA) to account for the multiplicity of potential models. Applying these methods to data from 105 U.S. cities for the period 1987-2005, we identify those cities having a high posterior probability of increased mortality risk during heat waves, examine the heterogeneity of the posterior distributions of mortality risk across cities, assess sensitivity of the results to the selection of prior distributions, and compare our BMA results to a model selection approach. Our results show that no single model best predicts risk across the majority of cities, and that for some cities heat wave risk estimation is sensitive to model choice. While model averaging leads to posterior distributions with increased variance as compared to statistical inference conditional on a model obtained through model selection, we find that the posterior mean of heat wave mortality risk is robust to accounting for model uncertainty over a broad class of models.

KEY WORDS: Climate change; Generalized additive models; Model uncertainty; Time series data

1. Introduction

Climate change is expected to cause an increase in the frequency, duration, and intensity of heat waves (Meehl and Tebaldi, 2004). An important component in assessing the impact of climate change on human health is quantifying the adverse health outcomes attributable to extreme weather events, such as heat waves (IWGCCH, 2010).

Many epidemiologic studies have investigated the health consequences of an extreme heat event, which has been selected for study in part because it has already exacted large health tolls. For example, it is estimated that over 700 people died in a single day as a result of the Chicago heat wave of 1995 (Semenza et al., 1996). Another extreme event, the European heat wave of 2003, led to 15,000 excess deaths in France (Fouillet et al., 2006), and additional thousands of deaths in other European countries (Garssen et al., 2005; Johnson et al., 2005; Conti et al., 2005; Grize et al., 2005). Additional single-site studies that have analyzed specific heat events have looked at Maricopa County, Arizona (Yip et al., 2008), Osaka Prefecture, Japan (Bai et al., 1995), St. Louis, Missouri (Bridger et al., 1976), and New York City (Ellis and Nelson, 1978), among others (Hertel et al., 2009; Rocklöv and Forsberg, 2009; Kovats et al., 2004).

The majority of heat wave studies have retrospectively analyzed the impact of specific, extreme heat events. While such studies are useful for identifying factors associated with increased risk of adverse health outcomes during heat waves, heat wave risks derived from these retrospective analyses are potentially biased because they tend to study only the highest impact heat events. Fewer studies have investigated the health effects of heat waves in multiple locations over time in order to quantify the risks attributable to less extreme, but still potentially harmful, heat wave events. In a study of three European cities, the estimated percent increase in mortality during heat wave days compared to non-heat wave days were 4.3% (95% CI 0.8 to 7.9) for London, 7.9% (3.6 to 12.4) for Budapest, and 15.2% (5.7 to 22.5) for Milan (Hajat et al., 2006). In another multi-site time series study of temperature-related mortality in the U.S., the pooled national estimate for increased risk associated with heat wave days, adjusted for temperature ranged from 3.2% (95% posterior interval 2.1 to 4.3) to 10.6% (6.1 to 15.3) across six heat wave definitions of varying levels of severity (Anderson and Bell, 2009).

Statistical models for estimating the relative risks of heat waves in the literature typically include both temperature terms as well as a heat wave day indicator. Under these models, the regression coefficient of the indicator provides an estimate of the "heat wave effect" beyond the effect of temperature. Since heat waves are functions of temperature, inclusion of both variables in the model may introduce multicollinearity. An alternative approach would be to simply not include temperature as a covariate in the regression model; however, since it is highly associated with mortality (Curriero et al., 2002; Hajat et al., 2002; Baccini et al., 2008; Basu, 2009), not including it yields a model with poor fit. Multiple temperature metrics have been considered both for inclusion in the models and as a basis for defining a heat wave event: daily maximum, minimum, or average daily temperatures, temperature measurements at various lags, as well as metrics that incorporate measures of humidity, such as humidex (Mastrangelo et al., 2007), apparent, and dew point temperature. The temperature terms have been modeled with a linear threshold model (Hajat et al., 2006), with cubic splines with fixed (typically 3 or 6) degrees of freedom (Anderson and Bell, 2009; Hajat et al., 2006), or with penalized splines where the degree of smoothing is determined by model selection on a smoothing parameter (Pauli and Rizzi, 2008). In the extensive air pollution epidemiology literature, temperature has been modeled with distributed lag models of both linear and nonlinear temperature covariates having time varying regression coefficients (Welty and Zeger, 2005). To estimate the effect of a heat wave event on mortality risk, studies have included day of week, smooth functions of relative humidity or dew point temperature, and smooth functions of calendar time to adjust for confounding (Anderson and Bell, 2009; Hajat et al., 2006; Pauli and Rizzi, 2008).

Studies investigating the shape of the temperature-mortality exposure-response relation have reported that the slopes of the exposure-response curves vary by geographical region (Curriero et al., 2002; Kalkstein and Davis, 1989). Previous multi-site studies of the association between heat waves and mortality have assumed the same regression model across locations, but they allow the exposure-response curves to vary across cities. In addition, while uncertainty in climate model predictions has been studied thoroughly, to our knowledge the degree of model uncertainty underlying heat wave risk estimates has not been systematically incorporated into risk estimation. Quantifying the risk of mortality and other health outcomes during heat waves, in conjunction with a comprehensive treatment of both statistical and model uncertainty, is a critical part of assessing the future impact of climate change on human health.

In this paper we attempt to overcome several challenges in estimating health risks associated with heat wave events. We develop an approach to estimate the relative risk of mortality associated with heat waves that flexibly models the nonlinear temperature-mortality association, incorporates multiple temperature metrics and their interactions, and allows the regression model to differ by city. More specifically, we first model the association between day-to-day changes in temperature covariates and mortality, with adjustment for potential confounders. Then we define and estimate the relative risk of mortality during heat wave days compared to non-heat wave days. Importantly, we implement Bayesian model averaging to systematically address model uncertainty. The outline of the paper is as follows. In section 2 we introduce the data and the definition of a heat wave event. Section 3 describes our within-city modeling approach and Section 4 details the model averaging implementation. In section 5 we apply our methods to a national analysis of heat wave mortality risk for 105

U.S. metropolitan areas, and we compare our results to a common model selection procedure. We summarize our findings and provide concluding remarks in section 6.

2. Data and Heat Wave Definition

We use data from the National Morbidity and Mortality Air Pollution Study (NMMAPS) (Samet et al., 2000). This dataset consists of daily time series of mortality counts, weather variables, and air pollution concentrations, from 1987 to 2005, in 105 of the largest U.S. urban communities. Mortality data were obtained from the National Center for Health Statistics (NCHS), and consist of counts of numbers of deaths classified by cause, stratified into three age categories: under 65, 65 to 74, and 75 and older. Daily weather data were obtained from the National Climatic Data Center, consisting of measures of daily average, minimum, maximum, and dew point temperature. Air pollution concentrations time series were provided by the US Environmental Protection Agency's Air Quality System. The data does not go beyond 2005 because the NCHS no longer makes available city-level daily mortality data.

Though there are many possible ways to define a heat wave that incorporate measures of intensity and duration, we use the definition introduced by the heat waves and climate change literature (Huth et al., 2000; Meehl and Tebaldi, 2004). This definition incorporates two temperature thresholds, namely the 97.5th percentile (T_1) and the 81st percentile (T_2) of daily maximum temperature `tmax`. We allow the thresholds T_1 and T_2 to vary by city, since notions of what constitutes extreme heat is different across U.S regions. A heat wave is then defined as the longest period of consecutive days satisfying the following three conditions: (1) The daily maximum temperature is above T_1 for at least 3 consecutive days; (2) the daily maximum temperature does not drop below T_2 during the entire period; and (3) the average of daily maximum temperature over the entire period is greater than T_1. For each city, we create an indicator variable $\texttt{hw}_t$ that is 1 if day t belongs to a heat wave event and

0 otherwise. Since our goal is to compare mortality during heat waves to mortality during non-heat wave periods, we restrict the data to include only the months May-October that contain the warm season. The primary outcome considered is all death mortality, excluding known accidental causes.

3. Modeling approach

For each city we model the association between day-to-day changes in temperature and mortality using a class of generalized additive models (Hastie and Tibshirani, 1990). We assume the daily number of deaths for the ith age category Y_{it} has a Poisson distribution with mean model

$$\log \mathbb{E}[Y_{it}] = \sum_{j=1}^{3} \gamma_i \texttt{age}_{it} + ns(t; \boldsymbol{\eta}, 3 \text{ df} \times 6 \text{ months}) + f(\texttt{temperature}_t; \boldsymbol{\beta}), \qquad (1)$$

where $f(\cdot; \boldsymbol{\beta})$ represents generically a series of one or more linear or nonlinear smooth functions of different temperature variables, `age` denotes an indicator for the age category (under 65, 65 to 74, and over 75), and $ns(t; \boldsymbol{\eta}, 3 \text{ df} \times 6 \text{ months})$ denotes natural cubic splines of time (indexed by t and parameterized by $\boldsymbol{\eta}$) with 3 degrees of freedom (df) per 6 months time and knots at quantiles.

[Table 1 about here.]

We consider a class of forms for the function $f(\cdot; \boldsymbol{\beta})$, which models the unknown temperature-mortality exposure-response relation. Table 1 summarizes the list of candidate models for each city. More specifically, the term $f(\cdot; \boldsymbol{\beta})$ may include one or more of the following variables: current day's maximum temperature `tmax`, average of the previous three days' maximum daily temperature $\overline{\texttt{tmax}}$,[3] and current day's dew point temperature `dptp`. These three weather variables were selected because each incorporates different features of temperature (current extreme, prior conditions, and humidity) that may pose health risks independently as main effects or jointly through interaction terms. We consider several models for the function $f(\cdot;\boldsymbol{\beta})$ in equation (1) that differ based on which of these three covariates ($\texttt{tmax}$, $\overline{\texttt{tmax}}^{(3)}$, $\texttt{dptp}$) are included and which of the possible interactions among them are incorporated into the model. The models also differ in the degree of nonlinearity they allow. Specifically, nonlinear functions of the temperature variables are modeled with natural cubic splines (denoted by $ns(\cdot)$), and multiple possible values for the degrees of freedom for the spline terms are considered. We define $g(\texttt{confounders}_{it};\boldsymbol{\gamma}) = \sum_{j=1}^{3}\gamma_i\texttt{age}_{it} + ns(t;\boldsymbol{\eta}, 3 \text{ df} \times 6 \text{ months})$, the functional adjustment for confounders, which is parameterized by $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \gamma_3, \boldsymbol{\eta})$. Inclusion of the smooth function of time accounts for seasonal and long-term trends in mortality. The form of the function $g(\cdot;\boldsymbol{\gamma})$ remains fixed across all models considered.

The relative risk of mortality associated with a heat wave day is defined as the average mortality on heat wave days divided by the average mortality on non-heat wave days, conditional on the confounders,

$$RR := \frac{\mathbb{E}(Y_{it} \mid \texttt{hw}_t = 1, \texttt{confounders}_t)}{\mathbb{E}(Y_{it} \mid \texttt{hw}_t = 0, \texttt{confounders}_t)} = \frac{\mathbb{E}[\exp\{f(\texttt{temperature}_t;\boldsymbol{\beta})\} \mid \texttt{hw}_t = 1]}{\mathbb{E}[\exp\{f(\texttt{temperature}_t;\boldsymbol{\beta})\} \mid \texttt{hw}_t = 0]}, \quad (2)$$

and we define $\theta = \log RR$ to be the log relative risk of mortality associated with a heat wave day. The equality in (2) is derived in Web Appendix A. Rather than selecting a specific model for f, we conduct Bayesian model averaging to calculate the posterior expectation averaged over a set of reasonable models for f (listed in Table 1). Details are discussed in the next section.

4. Bayesian model averaging

We first provide a brief overview of how BMA may be applied in the context of our problem and subsequently discuss prior specification and computational details.

The relative risk defined in equation (2) depends on the functional form of f through the parameters $\boldsymbol{\beta}$. Thus, the inferential target is the posterior distribution of the parameter $\boldsymbol{\beta}$

within each city. For a given city and a class of models $M_1, \ldots, M_K$, this is calculated by

$$\mathbb{P}(\boldsymbol{\beta} \mid \mathbf{y}) = \sum_{k=1}^{K} \mathbb{P}(\boldsymbol{\beta} \mid M_k, \mathbf{y})\mathbb{P}(M_k \mid \mathbf{y}), \tag{3}$$

which is a weighted average of the posterior distributions of the parameters that describe the unknown temperature-mortality exposure-reponse relation under each of the models, weighted by the models' posterior probabilities (Hoeting et al., 1999). The posterior probability of model M_k, is given by

$$\mathbb{P}(M_k \mid \mathbf{y}) = \frac{L(\mathbf{y} \mid M_k)\pi(M_k)}{\sum_{l=1}^{K} L(\mathbf{y} \mid M_l)\pi(M_l)}, \tag{4}$$

where

$$L(\mathbf{y} \mid M_k) = \int\int \mathbb{P}(\mathbf{y} \mid M_k, \boldsymbol{\beta}_k, \boldsymbol{\gamma}_k)\pi(\boldsymbol{\beta}_k, \boldsymbol{\gamma}_k \mid M_k)d\boldsymbol{\beta}_k d\boldsymbol{\gamma}_k \tag{5}$$

is the integrated likelihood of model M_k. Here $\boldsymbol{\beta}_k$ is the vector of parameters in the function $f_k(\texttt{temperature}; \boldsymbol{\beta}_k)$ and $\boldsymbol{\gamma}_k$ is the vector of parameters in the function $g_k(\texttt{confounders}_t; \boldsymbol{\gamma}_k)$ under M_k.

4.1 *Prior Specification*

There are several choices that go into computing $\mathbb{P}(M_k \mid \mathbf{y})$ using formulas (3)–(5). In particular, we must select (i) the class of candidate models $M_1, \ldots, M_K$, (ii) the models' prior probabilities $\pi(M_k)$, and (iii) the parameters' prior distributions $\pi(\boldsymbol{\beta}_k, \boldsymbol{\gamma}_k \mid M_k)$ for each model M_k. Our choice of the class of models relies mainly on using prior knowledge to construct a set of models to flexibly capture the temperature-mortality relation based on previous studies of this association. To allow the data to provide evidence as to which model is the most likely candidate for the data generating process, we place a discrete uniform prior on the candidate models.

In specifying a prior distribution on $(\boldsymbol{\beta}_k, \boldsymbol{\gamma}_k) \mid M_k$, again we would like a relatively non-informative prior to allow the data to "speak maximally". However, the prior specification must take into account nesting when models are nested within another (e.g. In Table 1, Models 1-6 are nested within Model 7). For example, if $\boldsymbol{\beta}_k = (b_1, \ldots, b_k)$ and $\boldsymbol{\beta}_{k+1} = (b_1, \ldots, b_k, b_{k+1})$, we want to ensure that $\pi(\boldsymbol{\beta}_{k+1} \mid b_{k+1} = 0) = \pi(\boldsymbol{\beta}_k)$. A reference class of prior distributions for generalized linear models that provides consistent information across nested models has been developed by Raftery (1996). This author found that this class of prior distributions depends on three parameter values, and the selection of only one of these (ϕ) has a significant impact on inference; thus, the author recommended to report results for a range of values of ϕ between 1 and 5 (Raftery, 1996). In our BMA implementation, we use this class of prior distributions with the parameter ϕ set to an intermediate value and consider other values in a sensitivity analysis. Web Appendix B gives further details of the prior distribution form and confirms, for our application, the insensitivity of results to the other two hyperparameters.

To provide a context for sensitivity analyses based on the parameter ϕ, we give some intuition on the implications of ϕ for the prior of the log relative risk θ (defined through equation (2)). In particular, it can be shown that, conditional on a model M_k, the variance of θ is approximately equal to ϕ^2 times a constant α_x that depends on the design matrix for model M_k. Since α_x may be computed for each model M_k, this relation may be used to incorporate prior information on θ through selection of ϕ (details in Web Appendix B).

4.2 *Computational Details*

In general, for generalized linear models there is no closed form solution to the integral in equation (5), and so we approximate the model posterior probabilities by using the Laplace approximation (Tierney and Kadane, 1986). To assess whether this approximation was reasonable, we conducted a simulation study (details in Web Appendix D), finding that the estimates of the posterior model probabilities based on the Laplace approximation performed well in our application. This yields estimates $\hat{p}_k$ of the probabilities $\mathbb{P}(M_k \mid \mathbf{y})$

for $k = 1, \dots, K$. We then obtain posterior samples of the log relative risk of mortality associated with a heat wave day θ through equation (3) as follows.

(1) Sample a model $M_k^{(j)}$ from its posterior distribution $\mathbb{P}(M_k \mid \mathbf{y})$, which is approximated by $\hat{p}_k$

(2) Sample $\boldsymbol{\beta}_k^{(j)}$ from $\mathbb{P}(\boldsymbol{\beta}_k, \mid M_k^{(j)}, \mathbf{y})$

(3) Compute

$$\theta^{(j)} = \log \frac{\frac{1}{n_1} \sum_t \exp\{f(\texttt{temperature}_t; \boldsymbol{\beta}_k^{(j)})\} \mathbf{I}(\texttt{hw}_t = 1)}{\frac{1}{n_0} \sum_t \exp\{f(\texttt{temperature}_t; \boldsymbol{\beta}_k^{(j)})\} \mathbf{I}(\texttt{hw}_t = 0)}, \tag{6}$$

where n_1 is the number of heat wave days and n_0 is the number of non-heat wave days in that city during the study period (May-Oct., 1987-2005).

Repeat this process for $j = 1, \dots, N = 2000$. To obtain the samples from $\mathbb{P}(\boldsymbol{\beta}_k, \mid M_k, \mathbf{y})$ in step 2, we implement the Metropolis-Hastings algorithm (Metropolis et al., 1953), where we check for convergence every 100 iterations (after a minimum of 800 iterations) based on the Monte Carlo standard error, and only keep the samples $\boldsymbol{\beta}_k$ after convergence has been achieved. Web Appendix C provides the details of our implementation. In this way, we obtain samples $\theta^{(j)}, j = 1, \dots, N$ from the posterior distribution of the log relative risk averaged over the class of candidate models, $\mathbb{P}(\theta \mid \mathbf{y})$.

4.3 *Comparison*

As an alternative to implementing BMA, a standard approach is to use a model selection criterion to determine the best fitting model and conduct inferences conditionally on this selected model. We consider the approach of fitting models $M_1, \dots, M_K$ from Table 1 and selecting the model M^* with the lowest Bayesian Information Criterion (BIC). We then compare the estimated posterior distribution of $\mathbb{P}(\theta \mid \mathbf{y})$ under BMA to the estimated posterior distribution $\mathbb{P}(\theta \mid M^*, \mathbf{y})$ under M^*, which is the BIC-selected model. Samples are obtained from $\mathbb{P}(\theta \mid M^*, \mathbf{y})$ by performing N iterations of steps 2–3 from section 4.2 using the same prior $\pi(\boldsymbol{\beta}_k, \boldsymbol{\gamma}_k \mid M^*)$ as was used in the full BMA. By comparing the posterior variance of θ estimated under BMA versus the posterior variance of θ under M^*, we assess the contribution of model uncertainty on statistical inference of heat wave risk.

5. Results

[Table 2 about here.]

Table 2 provides a summary of the distribution of heat wave events and the observed thresholds T_1 and T_2 for the twenty largest cities. Across the twenty largest cities, heat waves had the lowest frequency in Dallas/Fort Worth, Houston, and San Antonio at 0.6 heat wave events per year, and the highest frequency in Oakland, San Jose, and Seattle at 1.3 events per year. The city having the longest-lasting heat waves was Dallas/Fort Worth, where the average duration of a heat wave during the study period 1987-2005 was 19.4 days. Heat waves were of the shortest duration in San Jose and Oakland at 5.6 days per event. The 97.5th and 81st percentiles of maximum daily temperature (T_1 and T_2, respectively) were the lowest at 87.1°F and 75.2°F in Seattle and the highest at 111.9^{o}F and 105.1°F in Phoenix. While the cities with the largest number of heat wave days per year tend to be in the southern U.S. (Web Figure 1), there is not a clear latitude gradient; several cities in Florida and southern California have few to a moderate number of event days, as well many in the northeast and midwest.

[Figure 1 about here.]

Figure 1 displays bar plots summarizing the distribution of posterior model probabilities within each city. For the purpose of describing these results, we denote the cities as ordered from top to bottom in Figure 1 by $c_1, \ldots, c_{105}$. On the right side of the figure, in correspondence to each city, we denote the smallest number of models from the 33 model candidate set needed to contain 99% of the posterior mass. For several cities, a linear term of a single temperature predictor (Models 1-3) was the most likely data generating model, e.g. San Jose (c_{11}), Houston (c_{42}), and Miami (c_{82}); while for others spline terms of one or two temperature variables with multiple degrees of freedom (Models 8-9, 23-27) were more probable, e.g. New York (c_1), Philadelphia (c_{36}), and Detroit (c_{73}). Cities also varied based on the number of competing models that were plausible as the data generating mechanism. For several cities a single model contained the majority of the posterior mass, such as New York ($\hat{p}_{24}^{c_1} = 99.5\%$), Chicago ($\hat{p}_{27}^{c_2} = 98.8\%$), Jersey City ($\hat{p}_{8}^{c_3} = 98.1\%$), and Seattle ($\hat{p}_{8}^{c_4} = 97.9\%$). In other cases, two models shared the posterior mass nearly equally, as was the case for Atlanta (c_{45}), Phoenix (c_{60}), Oakland (c_{71}), and Detroit (c_{73}). In a few instances, namely Milwaulkee (c_{78}), Minneapolis/St. Paul (c_{79}), and Kansas City, KS (c_{105}), the distribution of posterior model probabilities was more diffusely spread out over the model space, requiring 10, 7, and 9, models, respectively, to contain over 99% of the posterior mass.

[Figure 2 about here.]

To capture the distinguishing features of the posterior distributions for each of the 105 cities, Figure 2 displays 95% highest posterior density (HPD) intervals for the log relative risk, where the grayscale of the intervals is proportional to kernel density estimates of the posterior of θ. We find that there is consistent evidence of an elevated risk of mortality during heat wave days in the industrial midwest, northeast, northwest, and southern California regions. In the upper midwest, there is some evidence of increased risk, though these posterior distributions are more diffuse. There is no consistent association in either the southeast or southwest. In addition, the degree of heterogeneity in the posterior mode varies by region, with the northeast exhibiting greater heterogeneity than the northwest and southeast. There is evidence of a bi- or multi-modal distribution for several cities, including Cleveland (region IM), Los Angeles (SC), Miami (SE), Phoenix (SW), and San Diego (SC). For other cities, such as Houston (SE) and San Antonio (SW), there is a spike in the posterior mass at zero, corresponding to no heat wave effect.

Of the 105 cities, 64 have posterior probability of $\theta > 0$ greater than 80% and 49 have posterior probability of $\theta > 0$ greater than 95%. To provide numerical results for the magnitude of the estimated heat wave effect, the posterior mean (95% HPD intervals) of the distribution of the percent increase in mortality on a heat wave day compared to a non-heat wave day for the three largest cities are 3.5% (2.3% to 4.6%) for Los Angeles, 10.6% (9.5% to 11.6%) for New York, and 12.4% (10.9% to 13.9%) for Chicago. There are 11 cities having over 95% posterior probability that the percent increase in mortality on a heat wave day is greater than 5%, namely Baltimore, Biddeford, Buffalo, Chicago, Detroit, Jersey City, New York, Providence, San Jose, Washington, and Worcester. The percent increase in mortality associated with a heat wave day is calculated as $100\% \times \{\exp(\theta^{(j)}) - 1\}$, where $\theta^{(j)}$ is a posterior sample of the log relative risk. Web Figure 2 shows a map of these results across the 105 U.S. cities. This map underscores the geographical pattern in heat wave mortality risk, with northern and west coast regions exhibiting heightened risk, and southeastern regions demonstrating little to no increased risk.

5.1 *Sensitivity analysis*

To assess the sensitivity of the posterior distribution of θ to prior specification, we considered a range of values for the prior dispersion parameter ϕ.

[Figure 3 about here.]

Figure 3 shows, for the twenty largest cities, kernel density estimates of the BMA posterior of the log relative risk θ under four different values of ϕ, and Web Table 1 shows corresponding numerical summaries across the 105 cities. We find that the parameter ϕ does not substantially impact the shape of the posterior for those cities with a clear cut "favorite" model, i.e. cities for which a single model contains nearly all of the posterior mass (e.g. Chicago, Dallas, New York). Among cities with a bi- or multi-modal posterior distribution of the log relative risk, some exhibit sensitivity to ϕ (Detroit, Los Angeles, and Phoenix) while others do not (Miami and Cleveland). In general, for cities where the likelihood of the models $M_1, \ldots, M_k$ is spread out over multiple models that differ in the number of df, the shape of the posterior distribution of θ will be sensitive to the prior parameter ϕ. Though different values of ϕ yield different weightings for the models describing heat wave mortality risk for certain cities, across these weighting schemes the magnitude of the risk as summarized by the posterior mean of θ remains relatively consistent (Web Table 1).

We additionally assessed sensitivity of our results to the class of models considered for BMA. We repeated our analysis with an expanded set of models that included the 33 from Table 1 as well as additional models from the literature that describe the temperature-mortality association. Specifically, we considered three models of the form of equation (1) from Anderson and Bell (2009), which have mean model

$$\log \mathbb{E}[Y_{it}] = \boldsymbol{\gamma}_1 \texttt{dow}_t + ns(t; \boldsymbol{\gamma}_2, 3 \text{ df} \times 6 \text{ months}) + ns(\texttt{tmax}_{t-lag}; \boldsymbol{\beta}_1, 3) + ns(D_t; \boldsymbol{\beta}_2, 3), \quad (7)$$

where $\boldsymbol{\gamma}_1$ is a vector of regression coefficients on the categorical variable for day of the week `dow`, $ns(\texttt{tmax}_{lag}; \boldsymbol{\beta}_1, 3)$ denotes natural cubic splines of maximum daily temperature for a given lag from day t with 3 df, and $ns(D_t; \boldsymbol{\beta}_2, 3)$ represents natural cubic splines of adjusted dew point temperature on day t, again with 3 df. As in Anderson and Bell (2009), we considered three lags for `tmax` (0, 1, and 2). This model is a slight adaptation of their model, because we used only data from half of the year containing the warm season, while they used a full year of data. As such, we necessarily reduced the number of degrees of freedom in the smooth function of time from 7 df to 3 per year. We found, consistent across the 105 cities, that these three models were not well supported by the data, having zero or nearly zero posterior mass.

5.2 *Comparison to a single model*

Model selection

Figure 3 shows, for the twenty largest cities, kernel density estimates of the posterior of the log relative risk θ under the BIC-selected model M^*, and Web Table 1 compares posterior mean and standard deviation estimates conditional on M^* to estimates derived from the BMA approach. For the majority of cities (94 out of 105), the model with the highest posterior probability is the same as the model that was selected under the BIC criterion. Comparing the estimate of $\mathbb{P}(\theta \mid \mathbf{y})$ under BMA for different values of ϕ to the estimated posterior under the BIC-selected model $\mathbb{P}(\theta \mid M^*, \mathbf{y})$ in Figure 3, we see that for some cities these posteriors are very similar, including Chicago, Dallas/Fort Worth, and San Jose. For other cities, such as Los Angeles, San Diego, and Miami, there is a noticeable discrepancy between estimates of the posterior distribution of the heat wave log relative risk under BMA versus BIC. In general, the BMA approach and the BIC model selection approach lead to divergent posteriors for those cities where two or more models provide similar fits to the data; in these situations, the posterior under BMA may be multi-modal or skewed, and it is typically more spread out than the posterior under the BIC-selected model. When we compute the sample standard deviation $\hat{\sigma}_{BMA}$ and $\hat{\sigma}_{BIC}$ of the BMA and BIC posterior samples of θ, respectively, we find that $\hat{\sigma}_{BMA} \geq \hat{\sigma}_{BIC}$ for nearly all of the cities (97 out of 105), and that on average $\hat{\sigma}_{BIC}$ is nearly 30% smaller than $\hat{\sigma}_{BMA}$ (Web Table 1).

Model from literature

Given that there is no “gold standard” model for estimating heat wave mortality risk, we also compare our results to model (7) with a lag of 0 on `tmax`, which is adapted from Anderson and Bell (2009). While the model selection approach based on the BIC criterion generally yields smaller posterior standard deviation estimates as compared to averaging over a set of multiple models, fitting this model from the literature across the 105 cities leads to larger posterior standard deviation estimates than does the BMA approach (see Web Table 1). As described above, when we included model (7) within the candidate set, this model was shown to be little supported by the data. As such, though heat wave risk estimates derived conditional on this model have larger uncertainty, this uncertainty is not incorporated into the BMA variability estimate.

6. Discussion

In this paper we develop a Bayesian model averaging approach to estimate the relative risk of mortality associated with heat waves. We apply this methodology in the most extensive study of heat wave mortality to date, covering 105 cities over 19 years of data (May 1987 - Dec. 2005).

Our proposed approach overcomes many of the challenges in estimating the adverse health effects of heat wave events. First, temperature variables are included in the model as predictors of mortality rather than for the purpose of adjustment. In other words, instead of viewing temperature and dew point temperature as confounding the heat wave-mortality association, these temperature variables are viewed as components of a heat wave that describe its various features. For example, higher current and previous days' temperature characterize more extreme heat wave events and have been shown to be important predictors of mortality (Basu, 2009). Thus, our goal is to estimate the total heat wave effect, defined as the expected number of deaths on heat wave days divided by the expected number of deaths on non-heat wave days during the warm season, adjusted for time-varying confounders such as season and long-term trends. Second, within each city, we specify a semi-parametric model to flexibly capture the nonlinear relation between several weather variables and mortality. The model makes as few assumptions as possible about the shape of the exposure-response function and does not require the cities to have the same model or even to include the same temperature predictors. This allows for heterogeneity of the temperature-mortality association across cities, in accordance with findings in prior studies that the shape of the temperature-mortality curve varies by U.S. region (Curriero et al., 2002). Third, we incorporate model uncertainty in the specification of the temperature-mortality exposure-response function by conducting Bayesian model averaging.

While BMA has been used in air pollution epidemiology studies (Thomas et al., 2007; Clyde, 2000; Clyde et al., 2000) and to evaluate stroke risk (Volinsky et al., 1997), among other risk assessment studies (Bailer et al., 2005), to our knowledge it has not been used in studies of temperature or heat waves and mortality. Rather, a primary model for the association of temperature variables and heat waves with the outcome is selected and other secondary models are considered through sensitivity analyses (Anderson and Bell, 2009; Hajat et al., 2006) or a model selection procedure is employed. In a previous study of heat waves and hospital admissions, a resampling procedure was proposed to assess the robustness of the model selection criterion (in this case the UBRE score) when a set of competing models, which may contain many similar models, is considered (Pauli and Rizzi, 2008).

Applying our approach to 105 U.S. cities, we found a heightened risk of mortality during heat wave days compared to non-heat wave days, especially in northern regions of the U.S. For example, we estimated a percent increase in mortality (95% HPD interval) of 8.8% (5.4% to 11.9%) in Washington, D.C., 3.7% (0.7% to 6.5%) in St. Louis, and 5.6% (3.9% to 7.6%) in Seattle. We generally did not find increased risk in southern regions. Though temperatures are higher at lower latitudes causing heat waves to be more extreme, it is possible that individuals and/or communities in hot climates have adopted precautions to limit the impact of heat, such as air conditioning use. For two cities (Oklahoma City and Shreveport, LA), we estimated a protective heat wave effect. Both had Model 2 as the most probable model of the set of 33 candidates, with temperature-mortality exposure-response function $f(\overline{\texttt{tmax}}^{(3)}; \boldsymbol{\beta}) = \beta_1 \overline{\texttt{tmax}},^{(3)}$ a linear term of the average of the previous three days' temperature. Several factors could contribute to a protective effect, including prevalence of air conditioning, presence of a heat response plan, and differences in demographic factors related to heat wave susceptibility. Previous studies have identified effect modifiers of the temperature-mortality association (Basu, 2009), but further research is needed to determine which factors explain heterogeneity across cities in heat wave mortality so that targeted interventions may be developed.

Implementing our approach for 105 cities and conducting BMA provides valuable insight into modeling heat wave mortality risk. First, we found that across cities the posterior distribution of model probabilities varies widely, and no single model best captures the temperature-mortality exposure-response function. Thus it is important for multi-site studies of heat wave risk to allow for heterogeneity in model specification across cities. Second, we found that for some cities heat wave risk estimation is sensitive to model selection, as demonstrated by the multi-modality of the posterior distributions under BMA. Additionally, our comparison of the posterior variance of the log relative risk under BMA to the variance under a single model emphasizes a twofold benefit of model averaging. On the one hand, models are weighted by their posterior probabilities so that the uncertainty of estimates conditional on less plausible models do not contribute to the model-averaged uncertainty estimate. On the other hand, when multiple models are plausible, BMA incorporates the variability from each potential model; thus conditioning inference on a single model obtained through a model selection procedure likely underestimates statistical uncertainty. In fact we found that the posterior standard deviation estimates of θ under the BIC-selected model were smaller compared with those under BMA. Nonetheless, our results demonstrate that the association of heat waves with elevated mortality is robust to accounting for model uncertainty over a broad class of candidate models and a range of prior distributions.

We accounted for model uncertainty in our analysis, but our results still depend on certain choices, such as the class of candidate models, the set of temperature covariates we considered for the function $f(\texttt{temperature}_t; \boldsymbol{\beta})$, the adjustment for confounders, and the choice of prior distributions for BMA. Any model selection procedure must similarly first determine a set of models to consider, and our BMA approach does not preclude the possibility that a better-fitting model exists outside of the model candidate set. Findings from previous studies informed the development of our candidate model set and the particular weather variables included. Time series generalized additive models for count data have been widely used in studies of the association of temperature and mortality (Basu and Samet, 2002; Basu, 2009). Additionally, models including smooth functions of current days' temperature together with smooth functions of an average of previous days' temperature and dew point temperature have been used in studies of the effect of air pollution on mortality, and have been shown to adequately control for weather effects (Welty and Zeger, 2005). Curriero et al. (2002) found that current days' temperature and the average of the previous three days' temperature had stronger associations with mortality than temperature variables at further lags, and that dew point temperature added predictive power to the model in a study of 11 U.S. cities. We included adjustment for confounders through a smooth function of time to account for seasonal and longer term trends in mortality, as has been used in many previous epidemiological studies of the effects of air pollution on health and previous temperature and heat waves time series studies (Peng et al., 2006; Bell et al., 2004). Future work might extend the set of candidate models in order to assess uncertainty in specifying the confounder adjustment model and to evaluate the importance of including PM, ozone, and other air pollution variables in the model. While none of the models we considered model serial dependence of the mortality counts, exploratory data analysis suggested that autocorrelation has little impact for this application. To assess the possibility of serial dependence in light of the different degrees of smoothing among the 33 candidate models, we fit a simple, intermediate, and complex model (models 2, 24, and 31) to each city's data and estimated the autocorrelation function (ACF) of the deviance residuals for each model and city combination. Visual inspection of the ACFs showed no consistent pattern, indicating that working independence is reasonable for our data. While we found that estimates of the posterior model probabilities based on the Laplace approximation were reasonable in this application, future work could explore the use of more accurate approximations such as integrated nested Laplace approximations (Rue et al., 2009).

We considered a discrete uniform prior on the set of candidate models, though it may be desired to include prior information on some of the models. Since we assumed a uniform prior, the posterior model probabilities calculated are proportional to the likelihood, and so Figure 1 may be examined to determine for which cities inclusion of prior information on the models would impact the results. In particular, we observe that under the uniform prior, over 85% of the posterior mass for cities $c_1 - c_{23}$ is contained in a single model, implying that including prior information on the models will not highly impact the results. On the other hand, cities having a likelihood that is spread out, such as $c_{78} - c_{105}$, will be more heavily impacted by the incorporation of prior information.

The sensitivity of our results to prior specification on the vector of regression parameters was assessed by considering a range of values of the dispersion hyperparameter ϕ. We also considered the addition of three models of the temperature-mortality association from the literature to the set of candidate models considered and found that their inclusion did not impact our results. We did not directly investigate the sensitivity of the choice of heat wave definition on the posterior distribution of the log relative risk, since our goal was to conduct an analysis of heat wave risk, conditional on a particular definition of a heat wave, that incorporates model uncertainty in the risk estimation. While changing the heat wave definition would almost surely impact the risk estimates themselves (e.g. greater intensity as measured through the thresholds T_1 and T_2 yielding larger estimates), it would not influence our findings on the contribution of model uncertainty: since the set of temperature-mortality models we considered did not include the heat wave indicator, the calculation of the posterior model probabilities in equation (4) is independent of the selected heat wave definition. Nonetheless, a better grasp of the features of heat waves that most impact mortality, such as intensity, duration, or timing in the summer is an important direction of future work.

Understanding the contributions of different sources of uncertainty is an integral part of a systematic assessment of future health risks under climate change. In order to combine estimates of present and historical relative risk of mortality associated with heat waves with output from climate simulation models based on various climate change scenarios, a measure of the corresponding uncertainty is desired. This measure should include both model uncertainty as well as statistical uncertainty conditional on a given model. In this study we provide the first comprehensive assessment of heat wave risk that incorporates model uncertainty.

Supplementary Materials

Web Appendices, Tables, and Figures referenced in Sections 3, 4.1, 4.2, and 5 are available under the Paper Information link at the Biometrics website
`http://www.biometrics.tibs.org`.

Acknowledgements

The project described was supported by Award Numbers T32ES012871 and R01ES012054 from the National Institute of Environmental Health Sciences and Award Numbers EPA R83622 and EPA RD83241701. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Environmental Health Sciences of the National Institutes of Health nor of the EPA.

References

Anderson, B. and M. Bell (2009). Weather-related mortality: How heat, cold, and heat waves affect mortality in the United States. *Epidemiology 20*, 205–213.

Baccini, M., A. Biggeri, G. Accetta, T. Kosatsky, K. Katsouyanni, A. Analitis, H. R. Anderson, L. Bisanti, D. D'Ippoliti, J. Danova, B. Forsberg, S. Medina, A. Paldy, D. Rabczenko, C. Schindler, and P. Michelozzi (2008). Heat effects on mortality in 15 European cities. *Epidemiology 19*(5), 711–9.

Bai, H., M. N. Islam, H. Kuroki, K. Honda, and C. Wakasugi (1995). [Deaths due to heat waves during the summer of 1994 in Osaka Prefecture, Japan]. *Nihon Hoigaku Zasshi 49*(4), 265–74.

Bailer, A. J., R. B. Noble, and M. W. Wheeler (2005). Model uncertainty and risk estimation for experimental studies of quantal responses. *Risk Anal 25*(2), 291–9.

Basu, R. (2009). High ambient temperature and mortality: a review of epidemiologic studies from 2001 to 2008. *Environmental health : a global access science source 8*, 40.

Basu, R. and J. Samet (2002). Relation between elevated ambient temperature and mortality: a review of the epidemiologic evidence. *Epidemiologic Reviews 24*, 190–202.

Bell, M. L., J. M. Samet, and F. Dominici (2004). Time-series studies of particulate matter. *Annual review of public health 25*, 247–80.

Bridger, C. A., F. P. Ellis, and H. L. Taylor (1976). Mortality in St. Louis, Missouri, during heat waves in 1936, 1953, 1954, 1955, and 1966. *Environ Res 12*(1), 38–48.

Clyde, M. (2000). Model uncertainty and health effect studies for particulate matter. *Environmetrics 11*(6), 745–763.

Clyde, M., P. Guttorp, and E. Sullivan (2000). Effects of ambient fine and coarse particles on mortality in Phoenix, Arizona. *Journal of Exposure Analysis and Environmental Epidemiology*.

Conti, S., P. Meli, G. Minelli, R. Solimini, V. Toccaceli, M. Vichi, C. Beltrano, and L. Perini (2005). Epidemiologic study of mortality during the summer 2003 heat wave in Italy. *Environ Res 98*(3), 390–9.

Curriero, F., K. Heiner, J. Samet, and S. Zeger (2002). Temperature and mortality in 11 cities of the eastern United States. *American Journal of Epidemiology 155*(1), 80–87.

Ellis, F. P. and F. Nelson (1978). Mortality in the elderly in a heat wave in New York City, August 1975. *Environ Res 15*(3), 504–12.

Fouillet, A., G. Rey, F. Laurent, G. Pavillon, S. Bellec, C. Guihenneuc-Jouyaux, J. Clavel, E. Jougla, and D. Hémon (2006). Excess mortality related to the August 2003 heat wave in France. *International archives of occupational and environmental health 80*(1), 16–24.

Garssen, J., C. Harmsen, and J. de Beer (2005). The effect of the summer 2003 heat wave on mortality in the Netherlands. *Euro Surveill 10*(7), 165–8.

Grize, L., A. Huss, O. Thommen, C. Schindler, and C. Braun-Fahrländer (2005). Heat wave 2003 and mortality in Switzerland. *Swiss medical weekly : official journal of the Swiss Society of Infectious Diseases, the Swiss Society of Internal Medicine, the Swiss Society of Pneumology 135*(13-14), 200–5.

Hajat, S., B. Armstrong, M. Baccini, A. Biggeri, L. Bisanti, A. Russo, A. Paldy, B. Menne, and T. Kosatsky (2006). Impact of high temperatures on mortality: Is there an added heat wave effect? *Epidemiology (Cambridge, Mass) 17*(6), 632–8.

Hajat, S., R. S. Kovats, R. W. Atkinson, and A. Haines (2002). Impact of hot temperatures on death in London: a time series approach. *Journal of epidemiology and community health 56*(5), 367–72.

Hastie, T. and R. Tibshirani (1990). *Generalized additive models*. Chapman and Hall.

Hertel, S., A. L. Tertre, K.-H. Jöckel, and B. Hoffmann (2009). Quantification of the heat wave effect on cause-specific mortality in Essen, Germany. *Eur J Epidemiol 24*(8), 407–14.

Hoeting, J. A., D. Madigan, A. E. Raftery, and C. T. Volinsky (1999). Bayesian model averaging: a tutorial. *Statist. Sci. 14*(4), 382–417.

Huth, R., J. Kyselý, and L. Pokorná (2000, Jan). A gcm simulation of heat waves, dry spells, and their relationships to circulation. *Climatic Change*.

IWGCCH (2010). A Human Health Perspective On Climate Change: A Report Outlining the Research Needs of the Human Health Effects of Climate Change.

Johnson, H., R. S. Kovats, G. McGregor, J. Stedman, M. Gibbs, and H. Walton (2005). The impact of the 2003 heat wave on daily mortality in England and Wales and the use of rapid weekly mortality estimates. *Eurosurveillance 10*(7), 168–71.

Kalkstein, L. and R. Davis (1989). Weather and human mortality: an evaluation of demographic and interregional responses in the United States. *Annals of the Association of American Geographers 79*(1), 44–64.

Kovats, R. S., S. Hajat, and P. Wilkinson (2004). Contrasting patterns of mortality and hospital admissions during hot weather and heat waves in Greater London, UK. *Occupational and Environmental Medicine 61*(11), 893–8.

Mastrangelo, G., U. Fedeli, C. Visentin, G. Milan, E. Fadda, and P. Spolaore (2007). Pattern and determinants of hospitalization during heat waves: an ecologic study. *BMC Public Health 7*(1), 200.

Meehl, G. and C. Tebaldi (2004). More intense, more frequent, and longer lasting heat waves in the 21st century. *Science 305*(5686), 994–997.

Metropolis, N., A. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller (1953). Equation of state calculations by fast computing machines. *Journal of Chemical Physics 21*(6), 1087–1092.

Pauli, F. and L. Rizzi (2008). Analysis of heat wave effects on health by using generalized additive model and bootstrap-based model selection. *Journal of the Royal Statistical Society: Series C (Applied Statistics) 57*(4), 473–485.

Peng, R., F. Dominici, and T. Louis (2006). Model choice in time series studies of air pollution and mortality. *J. R. Statist. Soc. A 162*(2), 179–203.

Raftery, A. (1996). Approximate Bayes factors and accounting for model uncertainty in generalised linear models. *Biometrika 83*(2), 251–266.

Rocklöv, J. and B. Forsberg (2009). Comparing approaches for studying the effects of climate extremes - a case study of hospital admissions in Sweden during an extremely warm summer. *Glob Health Action 2*.

Rue, H., S. Martino, and N. Chopin (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society. Series B 71*, 319–392.

Samet, J., S. Zeger, F. Dominici, and F. Curriero (2000). The National Morbidity, Mortality, and Air Pollution Study (NMMAPS). Part I. Methods and methodological issues. *Health Effects Institute, Boston*.

Semenza, J. C., C. H. Rubin, K. H. Falter, J. D. Selanikio, W. D. Flanders, H. L. Howe, and J. L. Wilhelm (1996). Heat-related deaths during the July 1995 heat wave in Chicago. *New England Journal of Medicine 335*(2), 84–90.

Thomas, D. C., M. Jerrett, N. Kuenzli, T. A. Louis, F. Dominici, S. Zeger, J. Schwarz, R. T. Burnett, D. Krewski, and D. Bates (2007). Bayesian model averaging in time-series studies of air pollution and mortality. *J Toxicol Environ Health Part A 70*(3-4), 311–5.

Tierney, L. and J. Kadane (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association 81*(393), 82–86.

Volinsky, C., D. Madigan, A. Raftery, and R. Kronmal (1997). Bayesian model averaging in proportional hazard models: Assessing the risk of a stroke. *Journal of the Royal Statistical Society. Series C (Applied Statistics) 46*(4), 433–448.

Welty, L. J. and S. L. Zeger (2005). Are the acute effects of particulate matter on mortality in the national morbidity, mortality, and air pollution study the result of inadequate control for weather and season? A sensitivity analysis using flexible distributed lag models. *American Journal of Epidemiology 162*(1), 80–8.

Yip, F. Y., W. D. Flanders, A. Wolkin, D. Engelthaler, W. Humble, A. Neri, L. Lewis, L. Backer, and C. Rubin (2008). The impact of excess heat events in Maricopa County, Arizona: 2000–2005. *Int J Biometeorol 52*(8), 765–72.

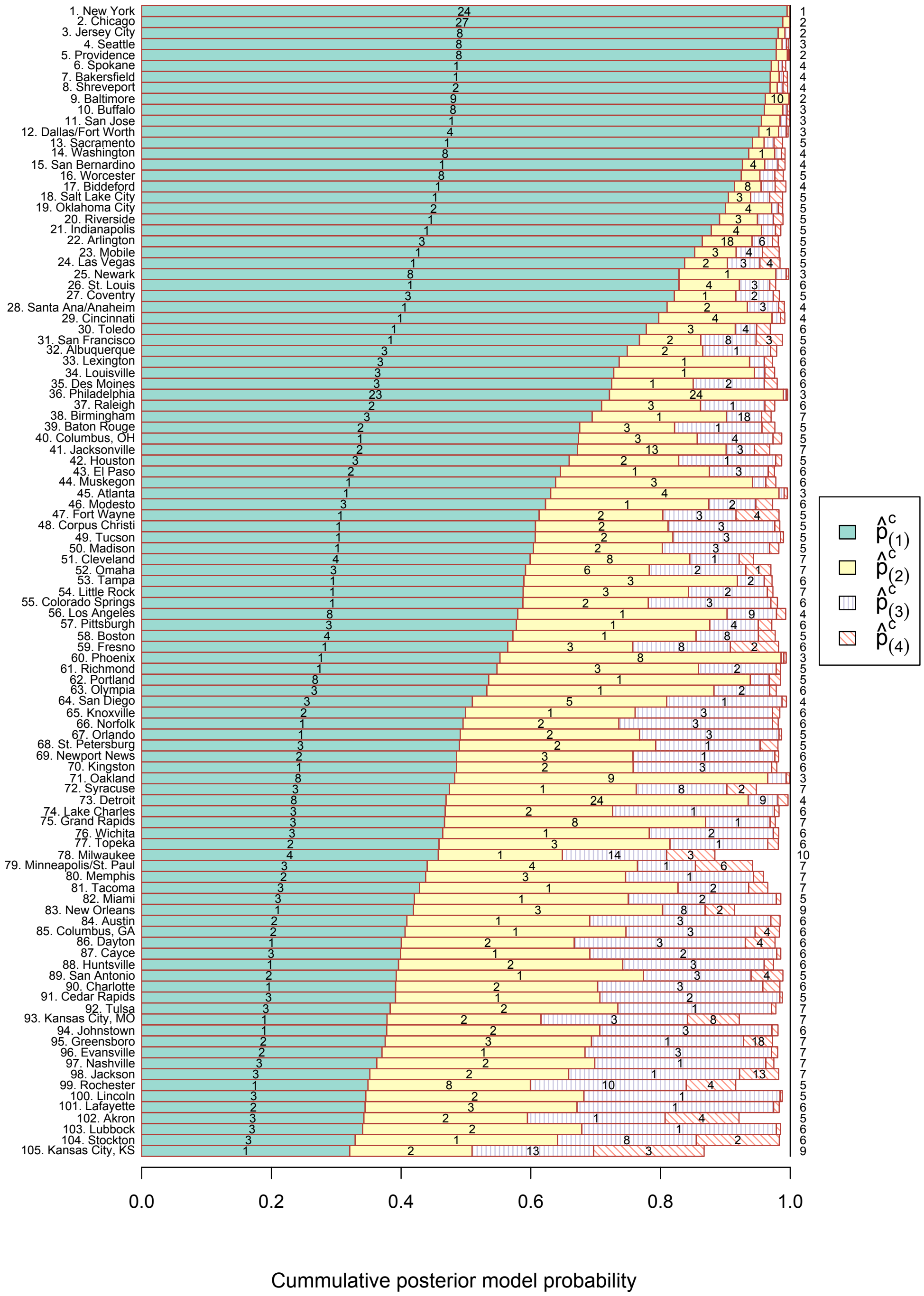

Figure 1. Bar plots summarizing the distribution of the posterior model probabilities $\hat{p}^c_k$ for the models M_k within each city c. Let $\hat{p}^c_{(k)}$ denote the kth largest estimated posterior probability for city c. For each city, the bar denotes the cumulative posterior probability $\sum_{k=1}^4 \hat{p}^c_{(k)}$. Number within each bar denotes which of the 33 models is represented by the bar; numbers at the right side of the plot denote the smallest number of models that contain 99% of the posterior mass, i.e. $\min\left\{K : \sum_{k=1}^K \hat{p}^c_{(k)} \geq 0.99\right\}$. Cities are listed from top to bottom in decreasing order of $\hat{p}^c_{(1)}$.

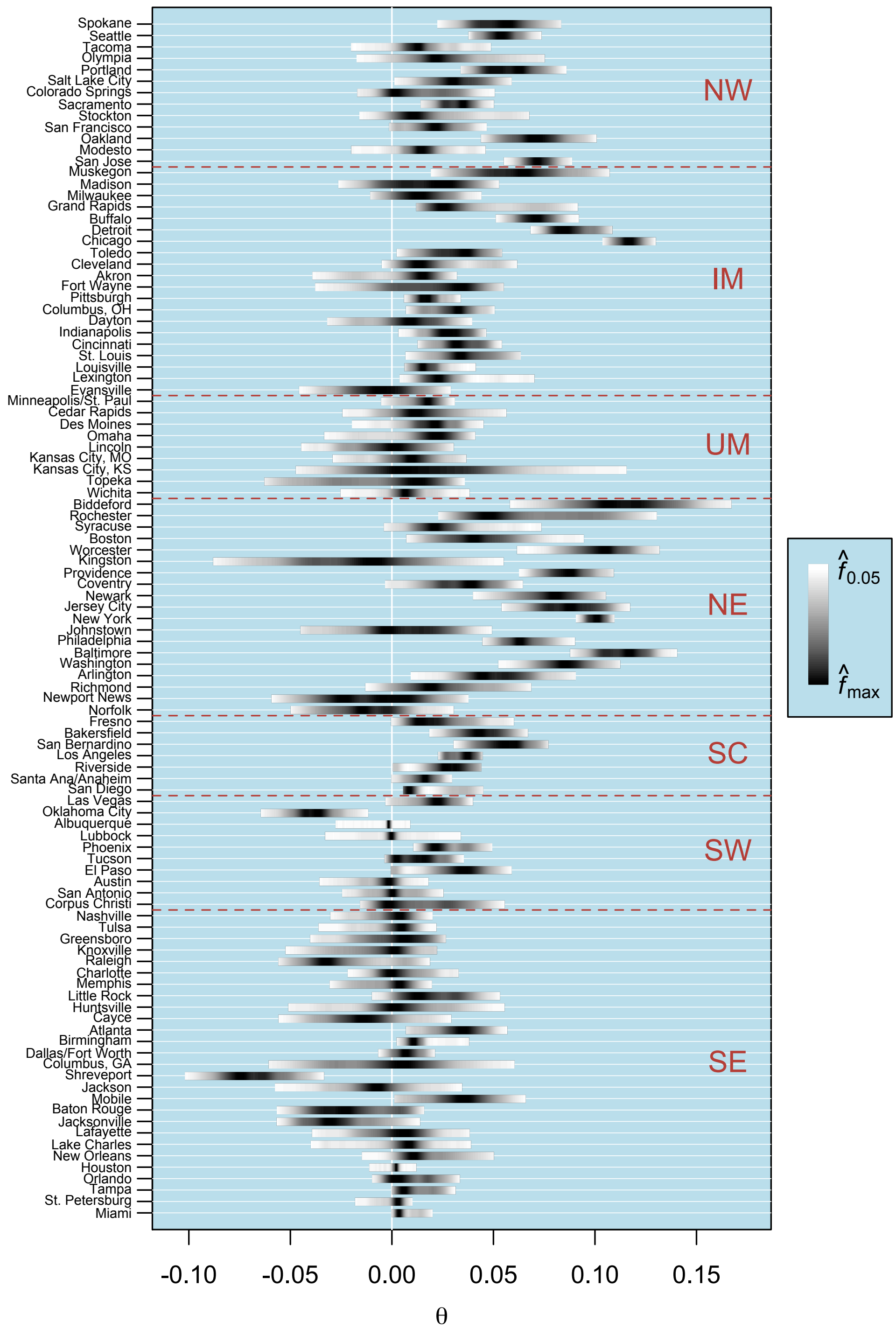

Figure 2. 95% highest posterior density (HPD) intervals for the log relative risk of mortality associated with a heat wave day θ. Shading is proportional to $\hat{f}$, the kernel density estimate of the posterior of θ, with black corresponding to $\hat{f}_{\max} = \max_\theta \hat{f}(\theta)$ and white corresponding to $\hat{f}_{0.05} = \min_{\theta \in H} \hat{f}(\theta)$, where H is the 95% HPD interval. Cities are categorized into 7 regions: southeast (SE), southwest (SW), southern California (SC), northeast (NE), upper midwest (UM), industrial midwest (IM), and northwest (NW). Within regions, cities are listed from top to bottom in order of decreasing latitude.

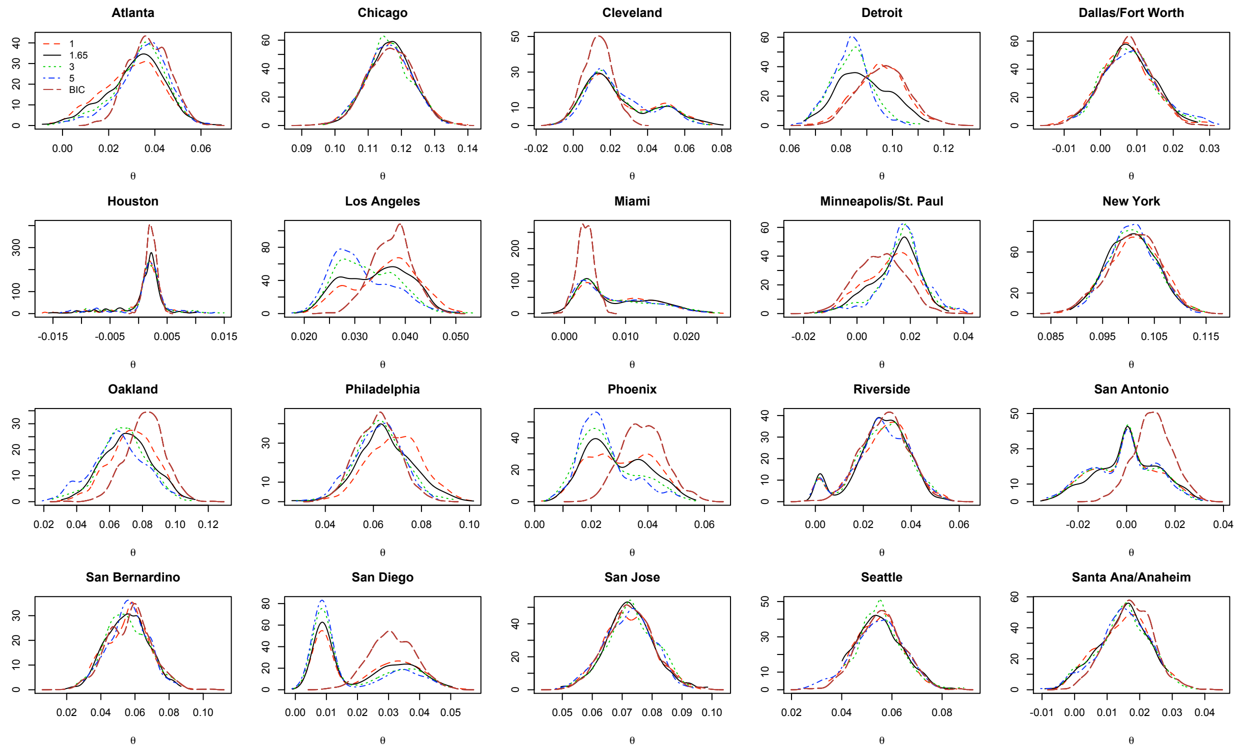

Figure 3. Kernel density estimates of the posterior $\mathbb{P}(\theta \mid \mathbf{y})$ under BMA for four values of the hyperparameter ϕ, as well as the kernel density estimate of the posterior under the BIC-selected model $\mathbb{P}(\theta \mid M^{*}, \mathbf{y})$ for the twenty largest cities, where θ is the log relative risk of mortality associated with a heat wave day. Legend is in the top left plot.

Models	$f(\texttt{temperature};\boldsymbol{\beta})$	df of natural cubic splines	Total df of $f(\cdot;\boldsymbol{\beta})$
1	$\beta_1\texttt{tmax}$		1
2	$\beta_1\overline{\texttt{tmax}}^{(3)}$		1
3	$\beta_1\texttt{dptp}$		1
4	$\beta_1\texttt{tmax}+\beta_2\overline{\texttt{tmax}}^{(3)}$		2
5	$\beta_1\texttt{tmax}+\beta_2\texttt{dptp}$		2
6	$\beta_1\overline{\texttt{tmax}}^{(3)}+\beta_2\texttt{dptp}$		2
7	$\beta_1\texttt{tmax}+\beta_2\overline{\texttt{tmax}}^{(3)}+\beta_3\texttt{dptp}$		3
8–12	$ns(\texttt{tmax};\boldsymbol{\beta},\lambda)$	$\lambda\in\{2,\dots,6\}$	2–6
13–17	$ns(\overline{\texttt{tmax}}^{(3)};\boldsymbol{\beta},\lambda)$	$\lambda\in\{2,\dots,6\}$	2–6
18–22	$ns(\texttt{dptp};\boldsymbol{\beta},\lambda)$	$\lambda\in\{2,\dots,6\}$	2–6
23–27	$ns(\texttt{tmax};\boldsymbol{\beta},\lambda)+ns(\overline{\texttt{tmax}}^{(3)};\boldsymbol{\beta},\lambda)$	$\lambda\in\{2,\dots,6\}$	4,6,8,10,12
28	$ns(\texttt{tmax};\boldsymbol{\beta},\lambda)+ns(\texttt{dptp};\boldsymbol{\beta},\lambda)$	$\lambda=3$	6
29	$ns(\overline{\texttt{tmax}}^{(3)};\boldsymbol{\beta},\lambda)+ns(\texttt{dptp};\boldsymbol{\beta},\lambda)$	$\lambda=3$	6
30	$ns(\texttt{tmax};\boldsymbol{\beta},\lambda)+ns(\overline{\texttt{tmax}}^{(3)};\boldsymbol{\beta},\lambda)+ns(\texttt{dptp};\boldsymbol{\beta},\lambda)$	$\lambda=3$	9
31	$ns(\texttt{tmax};\boldsymbol{\beta},\lambda)\times ns(\overline{\texttt{tmax}}^{(3)};\boldsymbol{\beta},\lambda)$	$\lambda=3$	15
32	$ns(\texttt{tmax};\boldsymbol{\beta},\lambda)\times ns(\texttt{dptp};\boldsymbol{\beta},\lambda)$	$\lambda=3$	15
33	$ns(\overline{\texttt{tmax}}^{(3)};\boldsymbol{\beta},\lambda)\times ns(\texttt{dptp};\boldsymbol{\beta},\lambda)$	$\lambda=3$	15

Table 1

Candidate statistical models for the relationship between daily temperature and log mortality. The first column contains the form of the function $f(\texttt{temperature};\boldsymbol{\beta})$*, where* `tmax` *is the maximum daily temperature,* $\overline{\texttt{tmax}}^{(3)}$ *is the average of the maximum daily temperature from the previous three days, and* `dptp` *is daily dew point temperature. The notation* $ns(\cdot;\boldsymbol{\beta},\lambda)$ *denotes a natural cubic spline with parameter vector* $\boldsymbol{\beta}$ *and degrees of freedom* λ*. The functional adjustment for confounders* $g(\texttt{confounders}_{it};\boldsymbol{\gamma})=\sum_{j=1}^{3}\gamma_i\texttt{age}_{it}+ns(t;\boldsymbol{\eta},3\ df\times 6\ months)$ *is the same across models.*

City	**Heat Wave**		**Thresholds (°F)**		**Average tmax (°F) during:**	
	Freq.	Duration	T_1	T_2	non-HW days	HW days
Atlanta	0.8	13.0	95.0	87.1	83.3	95.4
Chicago	0.7	10.4	93.0	82.9	76.8	94.0
Cleveland	0.9	9.7	91.0	82.0	75.9	91.9
Detroit	1.0	9.8	91.9	82.0	76.1	92.4
Dallas/Fort Worth	0.6	19.4	102.9	95.0	90.3	103.0
Houston	0.6	14.0	99.0	93.0	89.9	99.5
Los Angeles	1.2	9.0	104.0	95.0	90.1	104.4
Miami	0.7	12.4	93.9	91.0	89.2	94.3
Minneapolis/St. Paul	0.9	10.7	91.9	81.0	74.3	92.6
New York	0.8	7.7	96.1	84.9	79.9	96.6
Oakland	1.3	5.6	99.0	87.8	83.7	99.7
Philadelphia	1.1	10.8	95.0	86.0	80.8	95.5
Phoenix	0.7	12.8	111.9	105.1	100.2	112.4
Riverside	0.7	7.2	95.0	84.0	80.5	95.9
San Antonio	0.6	16.4	100.9	94.6	91.2	101.0
San Bernardino	0.7	6.4	89.6	79.0	76.5	91.0
San Diego	1.0	6.8	93.9	82.9	79.2	95.2
San Jose	1.3	5.6	99.0	87.8	83.8	99.7
Seattle	1.3	7.8	87.1	75.2	71.2	87.5
Santa Ana/Anaheim	1.2	8.1	98.1	87.1	82.6	98.5

Table 2

Heat wave and temperature summary statistics for the 20 largest cities, 1987-2005. Frequency is average number of heat waves events per year and duration is average length in number of days. Thresholds T_1 and T_2 are the 97.5th and 81st percentiles of maximum daily temperature tmax, respectively.