Barnett, Adrian G. and Tong, Shilu and Clements, Archie (2010) ***What measure of temperature is the best predictor of mortality?*** Environmental Research, 110(6). pp. 604-611.

What measure of temperature is the best predictor of mortality?

A.G.Barnett[*,a], S.Tong[a], A.C.A.Clements[b,c]

[a]School of Public Health & Institute of Health and Biomedical Innovation, Queensland University of Technology, 60 Musk Avenue, Kelvin Grove, Australia. Phone: +61 7 3138 6010, Fax: +61 7 3138 6010

[b]School of Population Health, The University of Queensland, Herston, Australia

[c]Australian Centre for Tropical and International Health, Queensland Institute of Medical Research, Herston, Australia

[*]Corresponding author

Email addresses: a.barnett@qut.edu.au (A.G.Barnett), s.tong@qut.edu.au (S.Tong), a.clements@uq.edu.au (A.C.A.Clements)

Abstract

Hot and cold temperatures significantly increase mortality rates around the world, but which measure of temperature is the best predictor of mortality is not known. We used mortality data from 107 US cities for the years 1987–2000 and examined the association between temperature and mortality using Poisson regression and modelled a non-linear temperature effect and a non-linear lag structure. We examined mean, minimum and maximum temperature with and without humidity, and apparent temperature and the Humidex. The best measure was defined as that with the minimum cross-validated residual. We found large differences in the best temperature measure between age groups, seasons and cities, and there was no one temperature measure that was superior to the others. The strong correlation between different measures of temperature means that, on average, they have the same predictive ability. The best temperature measure for new studies can be chosen based on practical concerns, such choosing the measure with the least amount of missing data.

Key words: climate, mortality, weather, temperature, apparent temperature, Humidex

Funding sources

Shilu Tong was supported by a National Health & Medical Research Council Research Fellowship (# 290515).

1. Introduction

Studies around the world have shown that daily mortality rates increase significantly with both hot and cold temperatures (Ballester et al., 1997; The Eurowinter Group, 1997; Keatinge et al., 2000; Donaldson et al., 2001; Braga et al., 2001; O'Neill et al., 2003; Goodman et al., 2004; Basu et al., 2005; Barnett et al., 2005; Schwartz, 2005; Analitis et al., 2008; Zanobetti and Schwartz, 2008; Rocklöv and Forsberg, 2008; Anderson and Bell, 2009). The risks depend strongly on climate and adaptation. People in colder climates tend to cope better in cold weather (The Eurowinter Group, 1997; Barnett et al., 2005), whilst those in hotter climates tend to cope better in extreme heat (Keatinge et al., 2000; Zanobetti and Schwartz, 2008).

Ambient air temperature may not be the best predictor of skin temperature, which is the main trigger of the body's cooling and warming mechanisms (Fanger, 1972; Ashcroft, 2000). Attempts have been made to combine temperature, humidity and wind to give a better estimate of the experienced temperature (Epstein and Moran, 2006). One alternative temperature measure is apparent temperature which combines mean temperature and dew-point temperature using the equation (Zanobetti and Schwartz, 2008):

$$\begin{aligned}\text{Apparent Temperature (deg F)} &= -2.653 + 0.994 \times \text{Mean Temperature (deg F)} \\ &\quad +0.0153 \times [\text{Dew-point temperature (deg F)}]^2 . \qquad (1)\end{aligned}$$

Maximum or minimum apparent temperature can be calculated using the same equation, with maximum or minimum temperature in place of mean temperature (Analitis et al., 2008). The aim of apparent temperature is to combine the effects of heat and cold with humidity. This measure has been used in a number of previous studies (Zanobetti and Schwartz, 2008; Baccini et al., 2008). For example, a rise in apparent temperature in the warm season was associated with increased all-cause mortality in adults using data from nine US cities (Zanobetti and Schwartz, 2008). Similarly, a rise in maximum apparent temperature in the warm season was associated with increased deaths in Europe (Baccini et al., 2008).

Another measure that attempts to combine temperature and humidity is the Humidex (Conti et al., 2005; Canadian Centre for Occupational Health and Safety, 2009), defined as

$$\begin{aligned} \text{Humidex (deg C)} &= \text{Mean Temperature (deg C)} + 0.5555\,(6.11E - 10)\,, \\ E &= \exp\left[5417.753\left(\frac{1}{273.16} - \frac{1}{\text{Dew Point Temperature (deg K)}}\right)\right], \end{aligned}$$

where deg K is degrees kelvin. The Humidex was designed by Canadian meteorologists to describe the feeling of hot and humid weather for an average person.

Temperature is strongly diurnal, and the range of temperatures during the day can be quite wide. This range is measured by the daily maximum and minimum temperature. The extremes of temperature will exert the most physiological pressure and so could be the most important predictor of mortality. Maximum temperature may also be a good measure of exposure because it often occurs in the middle of the day, which could coincide with a peak time for outdoor activity. Conversely, daily minimum temperatures are likely to occur at night when most people are in bed. In areas with good home insulation and heating the minimum temperature might therefore be a poor measure of actual exposure. The mean temperature, which summarises the entire day, may be a better estimate of exposure as it uses multiple observations per day and so should be less prone to measurement error compared with the temperature extremes. Many previous studies of mortality used average daily temperature (The Eurowinter Group, 1997; Keatinge et al., 2000; Braga et al., 2001), although others investigated the impact of minimum and maximum temperature (Schwartz, 2005).

Despite the great number of studies on the health effects of temperature, few studies have tried to objectively determine which measure of temperature is the best predictor of mortality. Metzger et al. (2009) examined mean, minimum and maximum temperature, the heat index and spatial synoptic classification, and chose the optimal combination of weather variables using the deviance and residual checks. Similarly, Hajat et al. (2006) compared mean, minimum, maximum and apparent temperature, and selected the best measure based on the deviance. In this paper we used cross-validation to pick the best model, as it gives more realistic predictions for future studies compared with statistics generated using the entire sample.

We aimed to find which temperature measure was the best predictor of mortality in order to better understand the mechanism of temperature-related mortality and to make recommendations for future studies. We compared seven temperature measures: mean, minimum and maximum temperature; mean, minimum and maximum apparent temperature; and the Humidex. We examined which measure gave the most accurate prediction of daily mortality. We used mean temperature based on the mean of minimum and maximum temperature because the mean temperature over 24 hours had large amounts of missing data. We assessed whether the predictive value of temperature depended on age, season and region, and whether including relative humidity gave better predictions. We also assessed whether there were any broad spatial patterns across the US in the best temperature measures.

2. Materials and methods

We used data from the National Morbidity and Mortality Air Pollution Study (NMMAPS) because it is publicly available, covers a wide range of climates, and has a large sample size (daily data for the years 1987–2000). The locations of the 107 cities used are shown in the supplementary figure S1. We excluded one city (Little Rock) because of a large amount of missing humidity data.

To summarise the correlations between the daily measures of temperature and humidity we calculated the Pearson correlations in each city and then averaged these correlations over the 107 cities.

2.1. Poisson regression model

We used Poisson regression with over-dispersion to model the association between temperature and daily counts of deaths. We selected Poisson regression as it is a common method for evaluating the association between temperature and mortality (Braga et al., 2001; Zanobetti and Schwartz, 2008). The Poisson model for the daily number of deaths on day d in each city is,

$$
\begin{aligned}
Y_d &\sim \mathrm{Po}(\mu_d), \\
\log(\mu_d) &= \boldsymbol{\alpha}\mathrm{dow}_d + \mathrm{ns}(d, \lambda_d) + \mathrm{ns}(\mathrm{Temperature}, \lambda_t, \lambda_l) \\
&\quad + \mathrm{ns}(\mathrm{Humidity}, \lambda_t, \lambda_l), \qquad d = 1, \ldots, 5114,
\end{aligned}
$$

where dow is a categorical term for day of the week (using a reference day of Sunday) and $\mathrm{ns}(., \lambda)$ refers to a natural spline with λ degrees of freedom (Ruppert et al., 2003). This term is used to control for secular trends and seasonal patterns in mortality, a greater λ_d means a greater flexibility, which means a stronger control for season. The temperature and humidity terms have the same two degrees of freedom: one for the temperature or humidity measure (λ_t) and the other for lag (λ_l), so that these effects are fitted using a non-linear surface (Armstrong, 2006). This surface is able to incorporate the non-linear U-shaped association between temperature and risk (with increases in risk at high and low temperatures), and the possibly non-linear association between exposure to temperature and a delayed (lagged) onset of death. We used a maximum delay of 25 days (Anderson and Bell, 2009). Because we did not know the best degrees of freedom, we fitted models for: 5, 6 and 7 degrees of freedom per year; 4, 5 and 6 degrees of freedom for temperature; and 3, 4 and 5 degrees of freedom for lag. We selected the best degrees of freedom from these 27 combinations using criteria described below.

We examined 14 different temperature and humidity models. We fitted each temperature measure without humidity: mean, minimum and maximum temperature; mean, minimum and maximum apparent temperature; and the Humidex. We combined mean, minimum and maximum temperature with mean humidity using the same natural spline basis as the temperature measure. In order to examine a simpler humidity effect we also used same day humidity with a natural spline using 3 degrees of freedom. As a "baseline" comparison we fitted a model without any measure of temperature or humidity, but with the terms for trend, season and day of the week. The models were fitted using the "dlnm" package in the R statistical software (Gasparrini and Armstrong, 2009).

To examine whether the best temperature measure differed by age we fitted separate models for the $<$ 65-year, 65–74-year and $\geq$ 75-year age groups.

2.2. Cross-validation

We calculated the predictive ability of each temperature measure using 10-fold cross-validation, which is a very robust model selection technique (Hen and Kamber, 2006). Cross-validation splits the sample into training and validation sets. The model is built using the training set and then tested using the validation set. This means that the inference is less tailored to the current data set, and cross-validation will give more realistic predictions for future studies (a key aim of our study).

To perform 10-fold cross-validation we randomly assigned a number between 1 and 10 to every day between 1 January 1987 and 31 December 2000. This numbering was done in random permuted blocks of 10 so that each number was equally represented and spread evenly over time. For each city and model we then ran the Poisson regression model 10 times, each time leaving out one of the 10 groups (10% of the data). The predicted values were then compared to the actual number of deaths to create the residuals. We used the squared Pearson residuals defined as (Dobson and Barnett, 2008)

$$r_{c,d} = (y_d - \hat{\mu}_d)^2 \Big/ \hat{\mu}_d \,, \qquad d \in D_c,\ c = 1, \ldots, 10,$$

where y_d is the observed number of deaths on day d, $\hat{\mu}_d$ is the estimated number of deaths, and D_c is the set of days left-out for cross-validation c.

We first averaged these daily residuals to give the mean for each left-out set:

$$\overline{r}_c = \sum_{d \in D_c} r_{c,d} \Big/ N_c \;,$$

where N_c is the number of days left-out in cross-validation c. This mean reflects the average difference between the observed and estimated number of deaths. The smaller this mean, the better the model.

We repeated the above cross-validation fives times, each time using a different set of randomly selected days. This was to ensure that our inferences were not overly-influenced by a particular random sampling pattern.

We verified the ability of this cross-validation method to find the best set of independent variables using a simulation study. In this simulation study we randomly created three independent variables, only one of which was associated with the randomly created dependent variable. We used a sample size of 5,114 days to match the NMMAPS data. We fitted seven different models (no independent variables, single independent variable, pairs of independent variables, all three independent variables) and found that the mean cross-validated residual was clearly lower for the correct model.

2.3. Summarising the model residuals

To examine the average performance of the models we used a regression model based on the mean residuals, $\overline{r}$. In the following definition we use subscripts for the mean residual from each model (a), city (b) and cross-validation (c). We modelled the mean residual using

$$\begin{aligned} \overline{r}_{a,b,c} &\sim \mathrm{N}(\mu_{a,b,c}, \sigma^2), \qquad a = 1, \dots, 24,\ b = 1, \dots, 107,\ c = 1, \dots, 50, \\ \mu_{a,b,c} &= \alpha + \beta_a + \gamma_{b,c}, \end{aligned}$$

where $\overline{r}_{a,b,c}$ is the observed mean and σ^2 is the estimated variance of the cross-validated residuals. The mean ($\mu_{a,b,c}$) was modelled using a linear regression equation with an overall intercept (α), a mean for each model (β_a), and a mean for each cross-validation within each city ($\gamma_{b,c}$). The regression parameters were given vague Normal prior distributions:

$$\begin{aligned}
\alpha &\sim \mathrm{N}(0, 10^5), \\
\beta_a &\sim \mathrm{N}(0, 10^5), \qquad a = 1, \dots, 24, \\
\gamma_{b,c} &\sim \mathrm{N}(\overline{\gamma}_b, 10^5), \qquad b = 1, \dots, 107,\ c = 1, \dots, 50, \\
\overline{\gamma}_b &\sim \mathrm{N}(0, 10^5), \qquad b = 1, \dots, 107.
\end{aligned}$$

We compared the performance of the models by plotting the estimated mean residual ($\hat{\alpha} + \hat{\beta}_a$) and its 95% credible interval. Separate estimates were made for each age group and in each season (and for all seasons combined). The seasons were defined as Winter (December, January, February), Spring (March, April, May), Summer (June, July, August) and Autumn (September, October, November). We also created estimates for seven US regions (Industrial Midwest, North East, North West, Other, Southern California, South East, South West, Upper Midwest). The cities in each region are shown in Table S1. To examine the variability in the best model in the same city we also estimated the mean cross-validated residual in each city and year.

A great advantage of using cross-validation is that the estimated mean residual ($\hat{\alpha} + \hat{\beta}_a$) will increase when a variable is added to the model that has no independent association with mortality. Standard likelihood based statistics, such as the deviance (Hajat et al., 2006; Metzger et al., 2009), always improve when a new variable is added to the model, even when that variable has no association with the dependent variable. This makes it difficult to assess the difference between models. A disadvantage of using the mean residual is that it is on an unfamiliar scale (being a squared and standardised residual), so as an additional measure of fit we used the Akaike information criterion (AIC, Akaike (1974)). However, we note that the AIC assumes an equal sample size, which is not always the true in the NMMAPS data because of some missing daily data for the temperature measures and humidity.

These models were fitted using a Bayesian paradigm (Dobson and Barnett, 2008).

We used the JAGS software to estimate the parameters (Plummer, 2008). We used a burn-in of 1,000 Markov chain Monte Carlo iterations and a sample of 1,000 subsequent iterations. We checked the convergence of the chains using the "coda" package in the R software (Plummer et al., 2009).

To examine geographical variation in the best predictors of mortality we selected the temperature measure in each city associated with the smallest average residual. The estimates from Anchorage and Honolulu were excluded from this part of the analysis as these cities are too distant from the contiguous United States. We used a support vector machine to find if there were regions where the same temperature measure generally gave the best predictions of mortality (Chang and Lin, 2008). We used the "e1071" package in R to make the estimates (Dimitriadou et al., 2009), and plotted the results using ArcView version 9.2 (ESRI, Redlands, CA). Additionally, we interpolated the Pearson residuals for each temperature measure using inverse distance weighting, again using ArcView. We then overlayed these interpolated values and selected the temperature measure that gave the minimum interpolated value at each location. We then created a map showing the best temperature measure for each location across the entire US.

3. Results

Table 1 shows the average correlations between the daily temperature measures and humidity. There were strong correlations (above 0.9) between most of the temperature measures. These strong correlations mean that we should expect only a small change in fit for different measures. Correlations between temperature and humidity were generally smaller, with the largest correlation between apparent temperature and humidity (correlation = 0.213).

Table 2 shows the mean cross-validated residual using data from all cities and seasons for the various degrees of freedom. In oldest age group the smallest mean residual used 5 degrees of freedom per year, 4 for temperature, and for 4 for lag. In the other two age categories the the smallest mean residual used 5 degrees of freedom per year, 4 for temperature, and for 3 for lag. The patterns in the AIC were similar to those for the residuals (Table S2), and for the AIC based on the 80 cities with less than 1% missing data (Table S3), and the mean cross-validated residual based on the 80 cities with less than 1% missing data (Table S4). The general pattern was an increased mean residual (and AIC) with increasing degrees of freedom for all three natural splines, so simpler models were favoured on average. From now on we show the results from each age group based on the combination of degrees of freedom that gave the smallest mean residual.

Figure 1 shows the mean cross-validated residuals for the 14 models split by the three age groups. The patterns in the < 65 and 65–74 year age groups were similar, as a model without any measure of temperature or humidity did best, the next best model used mean temperature, and the models including a surface for humidity did worst. In the ≥ 75 year age group the best models were apparent temperature and the Humidex; the models using a surface for humidity did poorly.

Figure 2 shows the mean cross-validated residuals in winter and summer for the three age groups. In the two youngest age groups a model without any measure of temperature or humidity did best. Mean temperature was the next best model in summer in these age groups. In the ≥ 75 year age group the best models in winter used minimum temperature or minimum apparent temperature, and the best model in summer used mean temperature. In all three age groups the six models including humidity were among some of the poorest fits in winter, but were more comparable to the other models in summer. Figure S2 shows the results in spring and autumn. For these two seasons a model without any measure of temperature or humidity did best in the youngest age group. In the ≥ 75 year age group the best model in spring used apparent temperature, and the best models in autumn used apparent temperature or the Humidex.

Figure 3 shows the mean cross-validated residuals in the ≥ 75 year age group by region (using all four seasons). There was great variability in the best model by region. In the Upper Midwest and South West a model without any measure of temperature or humidity did best. In the North West and South East the Humidex did best. In the North East the best model used maximum temperature, in the Industrial Midwest it was apparent temperature, and in Southern California mean temperature with humidity.

We examined the within-city variability by comparing the best model in each city and year (1987–2000) for the ≥ 75 age group. We defined the best model as that with the lowest mean cross-validated residual in each year. There were no cities where the same model had the lowest mean residual in all 14 years. The most consistent result was in Cayce, Iowa, where in 9 years out of 14 a model without any measure of temperature or humidity did best. In 100 cities the same best model was only selected in 6 years or fewer, indicating that there was little consistency in the best model.

We found no evidence of spatial variation in the best temperature measure in every season and age group. Many neighbouring cities had different best measures of temperature. When using support vector machines the error rates of the regions were high, as the predicted regions only correctly classified around 30% of cities. We relaxed the parameters of the support vector machine which increased the percent correctly classified to around 50%, but at the cost of producing multiple regions with no clear spatial pattern. Inverse distance weighting also produced multiple regions, indicating a lack of regional agreement. To reduce the influence of individual cities we used the ranking of the temperature models within a city rather than the mean residuals, but this still produced maps that were difficult to interpret and showed no smooth geographic variation in the best temperature measure. These results indicate that the best temperature measure was city-specific with little regional influence.

4. Discussion

Our results demonstrate that no temperature measure was consistently the best at predicting mortality in all age groups, seasons or regions. Instead we found marked variation in the best temperature measure across age groups (Figure 1), seasons (Figures 2 and S2) and regions (Figure 3). We also found variation in the best model between neighbouring cities, and there was no geographic consistency to the best models. We even found marked variation in the best model within cities in the $\geq$ 75 year age group, as there were no cities where the same model gave the best fit in every year.

The lack of consistency in the best model could be due to the relatively small risk of temperature-related mortality combined with the strong correlations in the temperature measures (Table 1). The relatively small risk of temperature-related mortality is evidenced by the results from the two youngest age groups where a model without any measure of temperature or humidity did best (Figures 1 and 2). In the oldest age groups there was a clear advantage to including most of the temperature measures, which reflects the greater risk posed by temperature in the frailest people. However, even in the oldest age group there was great variation in mean residuals by region (Figure 3) and there was no spatial consistency for the best models.

A few general patterns were more consistent. The models that included humidity as a separate variable tended to do worse, particularly when humidity was fitted using a natural spline surface with the same lag length and degrees of freedom as temperature. Models that included a simpler version of humidity by using the same day humidity with 3 degrees of freedom did better, particularly in summer (Figure 2). This suggests that effects of humidity are not as important or long-lasting as the effects of temperature. The effect of humidity on mortality should be secondary to the effect of temperature. Fanger estimated that a change in relative humidity of 0% to 100% can be compensated for by a change in temperature of only 1.5–3 degrees Celsius (Fanger, 1972). Increased humidity does make it more difficult to cool down in hot weather, as the body's evaporative cooling mechanism is compromised (Ashcroft, 2000). The loss of this cooling mechanism is potentially more serious in the elderly because of their increased frailty and reduced ability to thermoregulate (Horowitz and Robinson, 2007).

Despite not providing a clear recommendation of which temperature measure to use, there are some other useful messages in the results. We believe that the choice of the temperature measure is far less important that other modelling choices, such as the length of the lag and methods for dealing with non-linear risk (e.g., spline or polynomial) (Armstrong, 2006). We recommend choosing the temperature measure based on practical concerns, such as choosing the measure with the least amount of missing data, or, if temperature is available from a network of weather stations, choosing the measure that has the best spatial coverage of the study area. The similarity of the temperature measures also means that meta-analyses should not be too concerned about combining studies where different measures of temperature have been used (Bhaskaran et al., 2009), although between-study differences in lag lengths are likely to be important.

4.1. Limitations of the study

Whilst we tested a range of different degrees of freedom for the spline surfaces we did not test different bases functions, e.g., tensor products, radials (Ruppert et al., 2003), or methods that combine splines with a break-point for cold and hot effects (Muggeo, 2008). This could be an interesting area for future study.

We had no information on wind and so we could not examine the predictive value of a wind chill index. A strong wind can significantly reduce skin temperature, especially for people who go outside without adequate protective clothing. However, given the similarity in fit of the various models shown here we would be surprised if adding wind caused a big improvement in fit.

An alternative method for combining the effects of different aspects of the weather is the synoptic approach, where types of weather are classed into air mass groups (Gosling et al., 2009). We did not examine this method here, however it is difficult to imagine that a categorical model would do better than a model based on non-linear changes in risk. Samet et al. (1998) did compare two synoptic approaches with linear and non-linear regression methods, and found the synoptic approach gave a significantly poorer model fit for one synoptic approach and little difference with the other.

The NMMAPS data only contains three broad age groups, and the youngest age group (< 65 years) groups children with adults. The effects of temperature may be stronger in children than in adults (Gouveia et al., 2003), and there may also be differences in the lag if negative health effects occur faster in children than adults. Our model assumed a common shape for the effects of temperature, and hence the best measure of temperature needs further investigation in children.

We did not control for the effects of air pollution because we wanted to use the maximum amount of temperature information, and the air pollutant data are not available in all cities at all times in the NMMAPS study. Whilst it is known that air pollution and temperature have an interactive effect (Ren et al., 2006; Nawrot et al., 2007), there is also strong evidence for an independent effect of temperature on mortality (Welty and Zeger, 2005; Nawrot et al., 2007; Zanobetti and Schwartz, 2008). A recent study that also used the NMMAPS data concluded that there are, "separate and substantial mortality effects from temperature and from air pollution" (Anderson and Bell, 2009).

An important question for our results is how generalizable they are to other locations. We cannot be sure that these results are generalizable to other climates, although the NMMAPS data covers a wide range of climates. Similarly, we cannot be sure that these results are generalizable to other countries, because the interaction between people and the weather is modified by many factors, including housing conditions and clothing (Donaldson et al., 2001; The Eurowinter Group, 1997). Culture and adaptation to climate are also critical (Ashcroft, 2000).

Given the size of the data (covering cities, regions, seasons and age groups) it is possible that we have missed an important association between temperature and mortality in some smaller subgroup (e.g., winter in the $<$ 65-year age group in Southern California). So that interested readers can investigate such subgroups we have made the mean cross-validated residuals available in a supplementary data set.

4.2. Summary

We found large differences in the best temperature measure between age groups, seasons and cities, and there was no one temperature measure that was superior to the others. The strong correlation between different measures of temperature means that, on average, they have the same predictive ability. The best temperature measure for new studies can be chosen based on practical concerns, such choosing the measure with the least amount of missing data.

Acknowledgements

Thanks to the Department of Biostatistics at the Johns Hopkins Bloomberg School of Public Health and the Health Effects Institute for making the National Morbidity and Mortality Air Pollution Study data publicly available.

Thanks to the reviewers for their insightful and detailed comments.

Computational resources and services used in this work were provided by the High Performance Computer and Research Support Unit, Queensland University of Technology, Brisbane, Australia.

References

Akaike, H., 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control 19, 716–723.

Analitis, A., Katsouyanni, K., Biggeri, A., Baccini, M., Forsberg, B., Bisanti, L., Kirchmayer, U., Ballester, F., Cadum, E., Goodman, P. G., Hojs, A., Sunyer, J., Tiittanen, P., , Michelozzi, P., 2008. Effects of cold weather on mortality: Results from 15 European cities within the PHEWE project. Am J Epidemiol 168 (12), 1397–408.

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), 205–213.

Armstrong, B., 2006. Models for the relationship between ambient temperature and daily mortality. Epidemiology 17 (6), 624–631.

Ashcroft, F. M., 2000. Life at the Extremes: The Science of Survival. Flamingo, London.

Baccini, M., Biggeri, A., Accetta, G., et al, 2008. Heat effects on mortality in 15 European cities. Epidemiology 19 (5), 711–719.

Ballester, F., Corella, D., Pérez-Hoyos, S., Sáez, M., Hervás, A., 1997. Mortality as a function of temperature. A study in Valencia, Spain, 1991–1993. Int J Epidemiol 26 (3), 551–612.

Barnett, A. G., Dobson, A. J., McElduff, P., et al., 2005. Cold periods and coronary events: an analysis of populations worldwide. J Epidemiol Community Health 59 (7), 551–557.

Basu, R., Dominici, F., Samet, J. M., 2005. Temperature and mortality among the elderly in the United States. Epidemiology 16 (1), 58–66.

Bhaskaran, K., Hajat, S., Haines, A., Herrett, E., Wilkinson, P., Smeeth, L., 2009. Effects of ambient temperature on the incidence of myocardial infarction. Heart 95 (21), 1760–1769.

Braga, A. L. F., Zanobetti, A., Schwartz, J., 2001. The time course of weather-related deaths. Epidemiology 12 (6), 662–667.

Canadian Centre for Occupational Health and Safety, 2009. Humidex rating (http://www.ccohs.ca/oshanswers/phys_agents/humidex.html). (Accessed June 2009).

Chang, C.-C., Lin, C.-J., 2008. LIBSVM: a library for support vector machines (available at http://www.csie.ntu.edu.tw/∼cjlin/libsvm/). Tech. rep., Department of Computer Science and Information Engineering, National Taiwan University.

Conti, S., Meli, P., Minelli, G., Solimini, R., Toccaceli, V., Vichi, M., Beltrano, C., Perini, L., 2005. Epidemiologic study of mortality during the summer 2003 heat wave in italy. Environ Res 98, 390–399.

Dimitriadou, E., Hornik, K., Leisch, F., Meyer, D., , Weingessel, A., 2009. e1071: Misc Functions of the Department of Statistics (e1071), TU Wien. R package version 1.5-19.

Dobson, A. J., Barnett, A. G., 2008. An Introduction to Generalized Linear Models, 3rd Edition. Texts in Statistical Science. Chapman & Hall/CRC, Boca Raton, FL.

Donaldson, G. C., Rintamäki, H., Näyhä, S., 2001. Outdoor clothing: its relationship to geography, climate, behaviour and cold-related mortality in Europe. Int J Biometeorol 45 (1), 45–51.

Epstein, Y., Moran, D. S., 2006. Thermal comfort and the heat stress indices. Ind Health 44 (3), 388–398.

Fanger, P. O., 1972. Thermal Comfort: Analysis and Applications in Environmental Engineering. McGraw-Hill, New York.

Gasparrini, A., Armstrong, B., 2009. dlnm: Distributed Lag Non-linear Models. R package version 1.1.1. R package version 1.1.1.

Goodman, P. G., Clancy, L., Dockery, D. W., 2004. Cause-specific mortality and the extended effects of particulate pollution and temperature exposure. Environ Health Perspect 112 (2), 179–185.

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. Climatic Change 92 (3–4), 299–341.

Gouveia, N., Hajat, S., Armstrong, B., 2003. Socioeconomic differentials in the temperature-mortality relationship in sao paulo, brazil. Int J Epidemiol 32 (3), 390–397.

Hajat, S., Armstrong, B., Baccini, M., Biggeri, A., Bisanti, L., Russo, A., Paldy, A., Menne, B., , Kosatsky, T., 2006. Impact of high temperatures on mortality: Is there an added heat wave effect? Epidemiology 17 (6), 632–638.

Hen, J., Kamber, M., 2006. Data Mining: Concepts and Techniques, 2nd Edition. Elsevier.

Horowitz, M., Robinson, S. D., 2007. Heat shock proteins and the heat shock response during hyperthermia and its modulation by altered physiological conditions. In: Sharma, H. S. (Ed.), Neurobiology of Hyperthermia. Vol. 162 of Progress in Brain Research. Elsevier, pp. 433–446.

Keatinge, W. R., Donaldson, G. C., Cordioli, E., Martinelli, M., Kunst, A. E., Mackenbach, J. P., Nayha, S., Vuori, I., 2000. Heat related mortality in warm and cold regions of Europe: observational study. BMJ 321 (7262), 670–673.

Metzger, K. B., Ito, K., Matte, T. D., 2009. Summer heat and mortality in New York city: How hot is too hot? Environ Health Perspect 118 (1).

Muggeo, V. M., 2008. Modeling temperature effects on mortality: multiple segmented relationships with common break points. Biostatistics 9, 613–620.

Nawrot, T. S., Torfs, R., Fierens, F., 2007. Stronger associations between daily mortality and fine particulate air pollution in summer than in winter: evidence from a heavily polluted region in western Europe. J Epidemiol Community Health 61 (2), 146–149.

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.

Plummer, M., 2008. JAGS version 1.0.3 manual (available at http://www-fis.iarc.fr/∼martyn/software/jags/jags_user_manual.pdf). Tech. rep., International Agency for Research on Cancer.

Plummer, M., Best, N., Cowles, K., Vines, K., 2009. coda: Output analysis and diagnostics for MCMC. R package version 0.13-4.

Ren, C., Williams, G. M., Tong, S., 2006. Does particulate matter modify the association between temperature and cardiorespiratory diseases? Environ Health Perspect 114 (11), 1690–1696.

Rocklöv, J., Forsberg, B., 2008. The effect of temperature on mortality in Stockholm 1998–2003: a study of lag structures and heatwave effects. Scand J Public Health 36 (5), 516–23.

Ruppert, D., Wand, M. P., Carroll, R. J., 2003. Semiparametric regression. Cambridge University Press.

Samet, J., Zeger, S., Kelsall, J., Xu, J., Kalkstein, L., 1998. Does weather confound or modify the association of particulate air pollution with mortality? an analysis of the philadelphia data, 1973-1980. Environ Res 77 (1), 9–19.

Schwartz, J., 2005. Who is sensitive to extremes of temperature? A case-only analysis. Epidemiology 16 (1), 67–72.

The Eurowinter Group, 1997. Cold exposure and winter mortality from ischaemic heart disease, cerebrovascular disease, respiratory disease, and all causes in warm and cold regions of Europe. Lancet 349 (9062), 1341–1346.

Welty, L. J., Zeger, S. L., 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. Am J Epidemiol 162 (1), 80–88.

Zanobetti, A., Schwartz, J., 2008. Temperature and mortality in nine US cities. Epidemiology 19 (4), 563–570.

Figure 1. Mean cross-validated residuals (and 95% credible intervals) for the 14 different models by age group: $<$ 65 years (top row), 65–74 years (middle row), $\geq$ 75 years (bottom row). The smaller the residual, the better the model. App = Apparent; Hum^1 = Humidity with the same natural spline basis as temperature; Hum^2 = Same day humidity using a natural spline with 3 degrees of freedom.

Figure 2. Mean cross-validated residuals (and 95% credible intervals) for the 14 different models in winter (left column) and summer (right column) by age group: $<$ 65 years (top row), 65–74 years (middle row), $\geq$ 75 years (bottom row). The smaller the residual, the better the model. App = Apparent; Hum^1 = Humidity with the same natural spline basis as temperature; Hum^2 = Same day humidity using a natural spline with 3 degrees of freedom.

Figure 3. Mean cross-validated residuals (and 95% credible intervals) for the 14 different models in the $\geq$ 75 year age group by region. Top row: North West, Upper Midwest, North East. Middle row: Southern California, Industrial Midwest, South East. Bottom row: South West. The smaller the residual, the better the model. App = Apparent; Hum^1 = Humidity with the same natural spline basis as temperature; Hum^2 = Same day humidity using a natural spline with 3 degrees of freedom.

Supplementary data

Figure S1. The locations of the 105 cities in the contiguous United States from the National Morbidity and Mortality Air Pollution Study study. Anchorage and Honolulu were included in the study but are not shown on this map.

Figure S2. Mean cross-validated residuals (and 95% credible intervals) for the 14 different models in spring (left column) and autumn (right column) by age group: < 65 years (top row), 65–74 years (middle row), ≥ 75 years (bottom row). The smaller the residual, the better the model. App = Apparent; Hum^1 = Humidity with the same natural spline basis as temperature; Hum^2 = Same day humidity using a natural spline with 3 degrees of freedom.

Table 1: Average Pearson Correlations Between the Daily Temperature Measures and Humidity for the 107 US Cities, 1987–2000.

	Min	Max	AT	Min AT	Max AT	Humidex	RH
Mean temperature	0.963	0.971	0.955	0.933	0.962	0.988	−0.048
Min temperature		0.890	0.953	0.963	0.937	0.968	0.082
Max temperature			0.906	0.865	0.939	0.948	−0.146
AT				0.992	0.993	0.989	0.154
Min AT					0.975	0.973	0.213
Max AT						0.988	0.102
Humidex							0.051

Abbreviations: AT, Apparent Temperature; RH, Relative Humidity

Table 2: Mean cross-validated error by age group and the natural spline degrees of freedom: per year, for temperature and for lag. The smallest error in each age group is shown in bold.

DF per year	DF temperature	Age group < 65 DF lag = 3	< 65 DF lag = 4	< 65 DF lag = 5	65–74 DF lag = 3	65–74 DF lag = 4	65–74 DF lag = 5	≥ 75 DF lag = 3	≥ 75 DF lag = 4	≥ 75 DF lag = 5
5	4	**0.8367**	0.8375	0.8385	**0.7547**	0.7555	0.7562	0.9530	**0.9523**	0.9526
	5	0.8374	0.8385	0.8397	0.7552	0.7563	0.7572	0.9534	0.9530	0.9535
	6	0.8382	0.8396	0.8411	0.7559	0.7571	0.7583	0.9541	0.9540	0.9547
6	4	0.8397	0.8405	0.8414	0.7572	0.7580	0.7587	0.9537	0.9530	0.9533
	5	0.8404	0.8415	0.8427	0.7578	0.7588	0.7598	0.9542	0.9538	0.9543
	6	0.8412	0.8426	0.8441	0.7584	0.7597	0.7609	0.9549	0.9548	0.9556
7	4	0.8427	0.8436	0.8445	0.7599	0.7607	0.7614	0.9547	0.9541	0.9543
	5	0.8434	0.8446	0.8458	0.7605	0.7615	0.7625	0.9553	0.9549	0.9554
	6	0.8442	0.8457	0.8472	0.7611	0.7624	0.7636	0.9560	0.9559	0.9566

DF = degrees of freedom

Table S1: List of the 107 US cities by region.

Region	City	State	Region	City	State
Industrial Midwest	Akron	OH	Other	Anchorage	AK
($n = 20$)	Buffalo	NY	($n = 2$)	Honolulu	HI
	Chicago	IL	South East	Atlanta	GA
	Cincinnati	OH	($n = 26$)	Baton Rouge	LA
	Cleveland	OH		Birmingham	AL
	Columbus	OH		Cayce	SC
	Dayton	OH		Charlotte	NC
	Detroit	MI		Columbus	GA
	Evansville	IN		Dallas/Fort Worth	TX
	Fort Wayne	IN		Greensboro	NC
	Grand Rapids	MI		Houston	TX
	Indianapolis	IN		Huntsville	AL
	Lexington	KY		Jackson	MS
	Louisville	KY		Jacksonville	FL
	Madison	WI		Knoxville	TN
	Milwaukee	WI		Lafayette	LA
	Muskegon	MI		Lake Charles	LA
	Pittsburgh	PA		Memphis	TN
	St. Louis	MO		Miami	FL
	Toledo	OH		Mobile	AL
North East	Arlington	VA		Nashville	TN
($n = 19$)	Baltimore	MD		New Orleans	LA
	Biddeford	ME		Orlando	FL
	Boston	MA		Raleigh	NC
	Coventry	RI		Shreveport	LA
	Washington	DC		St. Petersburg	FL
	Jersey City	NJ		Tampa	FL
	Johnstown	PA		Tulsa	OK
	Kingston	NY	South West	Albuquerque	NM
	Newport News	VA	($n = 10$)	Austin	TX
	Norfolk	VA		Corpus Christi	TX
	Newark	NJ		El Paso	TX
	New York	NY		Las Vegas	NV
	Philadelphia	PA		Lubbock	TX
	Providence	RI		Oklahoma City	OK
	Richmond	VA		Phoenix	AZ
	Rochester	NY		San Antonio	TX
	Syracuse	NY		Tucson	AZ
	Worcester	MA	Southern California	Bakersfield	CA
North West	Colorado Spring	CO	($n = 7$)	Fresno	CA
($n = 14$)	Denver	CO		Los Angeles	CA
	Modesto	CA		Riverside	CA
	Oakland	CA		San Bernardino	CA
	Olympia	WA		San Diego	CA
	Portland	OR		Santa Ana/Anaheim	CA
	Sacramento	CA	Upper Midwest	Cedar Rapids	IA
	Salt Lake City	UT	($n = 9$)	Des Moines	IA
	San Francisco	CA		Kansas City	MO
	San Jose	CA		Kansas City	KS
	Seattle	WA		Lincoln	NE
	Spokane	WA		Minneapolis/St. Paul	MN
	Stockton	CA		Omaha	NE
	Tacoma	WA		Topeka	KS
				Wichita	KS

Table S2: Mean cross-validated AIC by age group and by the natural spline degrees of freedom per year, for temperature and for lag. The smallest AIC in each age group is shown in bold.

DF per year	DF tempe-rature	Age group: < 65			65–74			≥ 75		
		DF lag = 3	DF lag = 4	DF lag = 5	DF lag = 3	DF lag = 4	DF lag = 5	DF lag = 3	DF lag = 4	DF lag = 5
5	4	**17,215.0**	17,219.2	17,223.8	**16,542.6**	16,547.6	16,552.4	20,540.3	**20,537.4**	20,538.3
	5	17,218.5	17,224.1	17,230.0	16,546.4	16,552.7	16,558.9	20,542.3	20,540.5	20,542.5
	6	17,222.5	17,229.5	17,236.6	16,550.2	16,557.9	16,565.5	20,545.0	20,544.4	20,547.5
6	4	17,230.3	17,234.5	17,239.2	16,558.7	16,563.7	16,568.6	20,543.7	20,540.9	20,541.8
	5	17,233.9	17,239.5	17,245.4	16,562.7	16,569.1	16,575.2	20,546.1	20,544.4	20,546.4
	6	17,237.9	17,244.9	17,252.0	16,566.6	16,574.3	16,581.8	20,549.0	20,548.5	20,551.6
7	4	17,245.9	17,250.1	17,254.7	16,575.3	16,580.3	16,585.1	20,548.6	20,545.7	20,546.6
	5	17,249.5	17,255.0	17,260.9	16,579.2	16,585.5	16,591.7	20,551.2	20,549.4	20,551.4
	6	17,253.5	17,260.4	17,267.6	16,583.1	16,590.8	16,598.4	20,554.2	20,553.6	20,556.7

DF = degrees of freedom

Table S3: Mean cross-validated AIC by age group and the natural spline degrees of freedom per year, for temperature and for lag. Based on the 80 cities with less than 1% missing data for all temperature measures and humidity. Smallest AIC in each age group shown in bold.

DF per year	DF temperature	Age group								
		< 65			65–74			≥ 75		
		DF lag = 3	DF lag = 4	DF lag = 5	DF lag = 3	DF lag = 4	DF lag = 5	DF lag = 3	DF lag = 4	DF lag = 5
5	4	**18,214.1**	18,217.7	18,221.9	**17,442.6**	17,447.2	17,451.5	21,514.5	**21,509.9**	21,510.0
	5	18,217.3	18,222.2	18,227.7	17,446.1	17,451.9	17,457.5	21,515.9	21,512.3	21,513.4
	6	18,221.1	18,227.4	18,234.0	17,449.6	17,456.7	17,463.6	21,518.2	21,515.8	21,517.9
6	4	18,228.6	18,232.2	18,236.5	17,457.9	17,462.5	17,466.8	21,515.7	21,511.2	21,511.3
	5	18,232.0	18,236.8	18,242.3	17,461.6	17,467.4	17,473.0	21,517.6	21,514.1	21,515.2
	6	18,235.7	18,242.0	18,248.7	17,465.2	17,472.3	17,479.2	21,520.2	21,517.9	21,520.0
7	4	18,243.2	18,246.7	18,251.0	17,473.6	17,478.2	17,482.6	21,518.7	21,514.0	21,514.1
	5	18,246.5	18,251.4	18,256.9	17,477.3	17,483.1	17,488.7	21,520.8	21,517.2	21,518.2
	6	18,250.3	18,256.5	18,263.2	17,480.9	17,488.0	17,494.9	21,523.6	21,521.1	21,523.2

DF = degrees of freedom

Table S4: Mean cross-validated error by age group and by the natural spline degrees of freedom per year, for temperature and for lag. Based on the 80 cities with less than 1% missing data for all temperature measures and humidity. Smallest error in each age group shown in bold.

DF per year	DF temperature	Age group < 65 DF lag = 3	< 65 DF lag = 4	< 65 DF lag = 5	65–74 DF lag = 3	65–74 DF lag = 4	65–74 DF lag = 5	≥ 75 DF lag = 3	≥ 75 DF lag = 4	≥ 75 DF lag = 5
5	4	**0.8808**	0.8816	0.8825	**0.8088**	0.8096	0.8103	0.9712	**0.9702**	0.9703
	5	0.8814	0.8825	0.8838	0.8093	0.8104	0.8113	0.9715	0.9707	0.9710
	6	0.8822	0.8837	0.8852	0.8100	0.8113	0.8125	0.9720	0.9715	0.9720
6	4	0.8838	0.8845	0.8855	0.8114	0.8123	0.8129	0.9715	0.9705	0.9705
	5	0.8845	0.8855	0.8868	0.8120	0.8131	0.8140	0.9719	0.9711	0.9714
	6	0.8853	0.8867	0.8882	0.8127	0.8140	0.8152	0.9725	0.9720	0.9725
7	4	0.8869	0.8876	0.8886	0.8143	0.8152	0.8159	0.9721	0.9710	0.9711
	5	0.8876	0.8886	0.8899	0.8149	0.8160	0.8169	0.9726	0.9717	0.9720
	6	0.8884	0.8898	0.8913	0.8156	0.8169	0.8181	0.9732	0.9727	0.9732

DF = degrees of freedom