

Imperial College London
School of Public Health
Space and Atmospheric Physics

Climate change and health: statistical and stochastic modelling of
mortality and anomalous temperature stress events

Late Stage Review

Robbie Parks, PhD candidate

January 2018

Supervisors: Professor Majid Ezzati, Professor Ralf Toumi

Submitted in part fulfilment of the requirements for the degree of Doctor of Philosophy at Imperial
College London

Contents

1	Introduction	5
1.1	Relevance of climate change and health under a United Nations framework	5
1.2	Seasonal dynamics of mortality	5
1.3	Mortality under stress from anomalous temperature	5
1.4	Modelling anomalous temperature stress events and their impact on mortality	6
2	Aim and objectives	7
2.1	Aim	7
2.2	Objectives	7
3	Objective 1: Seasonal dynamics of cause-specific mortality in the USA	8
3.1	Background	8
3.2	Data	8
3.2.1	Mortality data	8
3.2.2	Temperature data	9
3.3	Methods	9
3.4	Results	10
3.4.1	Wavelet analysis	10
3.4.2	Centre of gravity analysis	10
3.4.3	Change in percentage difference in death rates between maximum and minimum mortality	11
3.4.4	Sub-national centre of gravity analysis	11
3.5	Discussion	12
4	Objective 2: Impact of large-scale ambient temperature changes on cause-specific mortality by age-sex group in the USA	14
4.1	Background	14
4.2	Data	14
4.3	Methods	15
4.4	Generation of temperature statistics	15
4.4.1	Statistical analysis	17
4.5	Results	18
4.6	Discussion	18
5	Objective 3: Risk assessment of large-scale ambient temperature changes using climate projections and IMAGE	19

5.1	Background	19
5.2	Data	19
5.3	Methods	19
5.4	Results	19
5.5	Discussion	19
6	Future work	20

List of Figures

List of Tables

1	ICD9 and ICD10 lookups for causes of deaths used in this study	8
2	Names of temperature statistics, their descriptions and why they are of interest. All centred statistics are based around a state-month long-term normal (°C), calculated from 1980-2009	16

1 Introduction

1.1 Relevance of climate change and health under a United Nations framework

It is incumbent upon the scientific and health community to work together to analyse patterns and trends in mortality due to changing weather patterns, and to provide means for political institutions to understand and act upon the findings.(9-11) The need to understand and manage a changing climate to preserve and improve human health is a key requirement of United Nations Sustainable Development Goal 3.D. Identifying the risk to communities is crucial for fulfilling this goal.(12)

1.2 Seasonal dynamics of mortality

It is well-established that death rates vary throughout the year, and in temperate climates there tend to be more deaths overall in winter than in summer.(3-6) Therefore, it has been hypothesized that a warmer world may lower winter mortality in temperate climates.(1, 2) In a large country like the USA, which possesses distinct climate regions, not only do average annual death rates vary geographically, but also the seasonality of mortality may vary, due to both localized weather patterns and regional differences in adaptation measures such as heating, air conditioning, and healthcare.(7-10) Different causes of death also possess distinct seasonal dynamics.[REF]

The presence and extent of seasonal variation in mortality may also itself change over time, due to shifts in weather regimes, lifestyle, adaptation technologies, and healthcare.(11-13) A thorough understanding of the long-term dynamics of seasonality of mortality by cause of death, and its geographical and demographic patterns, is needed to identify at-risk groups, plan responses at the present time as well as under changing climate conditions. There is however limited data to characterize the seasonality of mortality in relation to cause of death, age, sex, and local climate, or to understand how it has changed over time.

1.3 Mortality under stress from anomalous temperature

Each year, deaths occur from exposure to anomalous temperature patterns.(1) Such patterns include not only extreme heat and cold stress events, but also deviations from average long-term ambient temperature.(2) The Intergovernmental Panel on Climate Change (IPCC) has predicted with high confidence that the Earth's changing climate will result in a long-term paradigm shift in weather patterns, including higher average temperatures, along with more varied weather episodes.(3,4) Changes in anomalous heat event characteristics have also been observed in the USA, with statistically significant increases in heat events.(57) More common and more severe winter weather events in the USA are also possible.(8) As such, the response of human mortality due to weather patterns is set to evolve and potentially amplify due to forthcoming climate change.

Temperature patterns for every country will change under climate change.(13) Although exposure to changing temperature patterns is a global-scale phenomenon, vulnerability of a community will be affected by local socioeconomic, political, and geographical factors. As such, within even a single country, it is of importance to provide analysis of risk to different age-sex groups. This may be especially true for a country like the United States, which contains several distinct climates, as well as a variable political landscape, and where future population exposure to heat extremes is forecast to increase up to 6-fold when comparing the end of the 21st century to the 20th century base level.(14)

1.4 Modelling anomalous temperature stress events and their impact on mortality

There exists an ensemble of physics-based models, known as Global Climate Models (GCMs), produced by Regional Climate Centers (RCCs) throughout the world.[REF] These GCMs project future global climatic conditions under several plausible Representative Concentration Pathways (RCPs). Due to the computing expense of produce the GCMs, it is possible to project forward by a limited number of years. This results in being able to compute anomalous heat for return periods defined by the length of the time series, here up to 2100. [REF]

Stochastic weather generators can use meteorological data from the past or projections of the future as an input to generate much longer synthetic time series of weather.[REF] Such time series of arbitrary length can produce realistic time series of various weather metrics while preserving the spatial and temporal statistics of the input data. One state-of-the-art example of a stochastic weather generator is the Imperial College Weather Generator (IMAGE). The main advantage of using a stochastic weather generator like IMAGE is its ability to generate heat stress events at scales larger than the input data.

With a framework to understand the relationship between mortality and heat stress, outputs on returns periods from IMAGE can be used to make risk assessments of potential anomalous temperature stress events. These will be of value to decision-makers on a national and sub-national level for both planning for and mitigating the most harmful effects of climate change and extreme weather on health.

2 Aim and objectives

2.1 Aim

The overall aim of my thesis is to establish a logical framework of identifying and describing seasonality of monthly mortality, to implement a realistic model of monthly mortality which successfully captures its response to measures of heat stress, to attribute historical mortality to anomalous heat stress, and to estimate the potential effect of climate change on monthly mortality.

2.2 Objectives

My thesis research has three primary objectives that will together help achieve this aim:

1. To develop a method of identifying and describing seasonal dynamics of mortality.
2. To develop a mathematical model which can establish how monthly mortality is associated with measures of heat stress.
3. To estimate the effect of climate change scenarios on mortality by cause of death, gender, age, and geography, using both future scenario climate model data and/or data generated from stochastic weather generators such as the Imperial College Weather Generator (IMAGE).[REF]

3 Objective 1: Seasonal dynamics of cause-specific mortality in the USA

3.1 Background

Here, we comprehensively characterize the demographic, spatial and temporal patterns of cause-specific mortality seasonality in the entire USA, through the application of wavelet analytical techniques that have been used to study the dynamics of weather phenomena(14) and infectious diseases(15) to over three decades of national mortality data.

The work for objective 1 in chapter 3 is currently submitted and under review for publication.

3.2 Data

3.2.1 Mortality data

We used data on all 77,771,264 deaths in the USA from 1980 to 2013 from the National Center for Health Statistics (NCHS). Age, sex, state of residence, and month of death were available for each record. Yearly population counts were available from NCHS for 1990 to 2013 and from the US Census Bureau prior to 1990.(27) We inferred monthly population counts through linear interpolation, assigning each yearly count to July. We also subdivided the national data geographically by climate regions used by the National Oceanic and Atmospheric Administration (Figure XX).(28)

We also divided up the the deaths into 4 broad groups: cancer, cardiopulmonary, external, and other deaths. As the period of our study crossed between coding methods for the coding of death by the International Classification of Diseases (ICD) from ICD9 to ICD10 (with ICD10 used from 1999 to the present day), it was necessary to create a look-up table for equivalent causes of death from ICD9 to ICD10 (Table 1), with ‘Other’ being all codes not collected in Cancer, Cardiopulmonary or External.[ref]

Table 1: ICD9 and ICD10 lookups for causes of deaths used in this study

Cause of death	ICD9 coding	ICD10 coding
Cancer	140.0 - 239.9	C00 - D48
Cardiopulmonary	390.0 - 519.9	I00 - J99
External	800.0 - 999.9	S00 - Z99

Data were divided by sex and age in the following 10 age groups: 0-4, 5-14, 15-24, 25-34, 35-44, 45-54, 55-64, 65-74, 75-84, 85+ years. We calculated monthly death rates for each age and sex group, both nationally and for sub-national climate regions. Death rate calculations accounted for varying

length of months, by multiplying each months death count by a factor that would make it equivalent to a 31-day month.

3.2.2 Temperature data

We obtained data on temperature from ERA-Interim, which combines predictions from a physical model with ground-based and satellite measurements.(29) We used gridded four-times-daily estimates at a resolution of 80km to generate monthly population-weighted temperature by climate region throughout the analysis period.

3.3 Methods

For each cause of death, as well as for all-cause mortality, we used wavelet analysis to investigate seasonality, both nationally and sub-nationally, for each age-sex group. Wavelet analysis uncovers the presence, and frequency, of repeated maxima and minima in each age-sex-specific death rate time series. In brief, a Morlet wavelet, described in detail elsewhere,(30) is equivalent to using a moving window on the death rate time series and analyzing periodicity in each window using a short-form Fourier transform, hence generating a dynamic spectral analysis. The resulting coefficients can be presented on a two-dimensional plot of time against frequency (Figure XX). Importantly, wavelet analysis is able to measure dynamic seasonal behavior, in which the periodicity of death rates may disappear, emerge, or change over time. This is not possible in standard Fourier analysis or when fitting a statistical model with a period basis function. We used the R package WaveletComp (version 1.0) for the wavelet analysis. Before analysis, we logarithmically transformed death rates, detrended using a polynomial regression, and rescaled each all-cause mortality death rate time series so as to range between 1 and -1.

We identified age-sex groups whose wavelet power spectra differed from that of a white noise spectrum, which represents random fluctuations, at 5% significance level, for the entire study period (1980-2013). We then calculated the centre of gravity and the negative centre of gravity of monthly death rates. These parameters estimate when in the year, on average, maximum and minimum death rates occur, respectively. For calculating centre of gravity, each month was weighted by its death rate; for negative centre of gravity, each month was weighted by the difference between its death rate and the years maximum death rate. In taking the weighted average, we allowed January (month 1) to neighbour December (month 12), a technique known as circular statistics. Along with each circular mean, a 95% confidence interval (CI) was calculated by using 1000 bootstrap samples. The R package CircStats (version 0.2.4) was used for this purpose.

For each age-sex group and year, we used a Poisson model to estimate the percentage difference in death rates between the maximum and minimum mortality months for each year, and its standard

error which accounts for population size. We then fitted a linear regression to the time series of seasonal differences for each age and sex group, weighting each by the inverse of the square of its standard error. We calculated change in the fitted values from 1980 to 2013, reported as percentage point difference, as a quantitative measure of how the seasonality of death rates has changed over time.

3.4 Results

3.4.1 Wavelet analysis

All-cause mortality in males had a statistically significant 12-month seasonality in all age groups, except in ages 35-44 years, where it displayed statistically significant periodicity at 6 months (Figure XX). In females, there was no significant 12-month seasonality in ages 5 to 34 years (Figure XX); however, girls aged 5-14 years exhibited periodicity at 6 months for most of the analysis period. While seasonality persisted throughout the entire analysis period in older ages, it largely disappeared after late 1990s in children aged 0-4 years in both sexes and in women aged 15-24 years.

In males and females, Cancer mortality was statistically significant from ≥ 55 years (Figure XX). Cardiopulmonary mortality was statistically significant for all ages for both males and females, though is relatively across time for ages 5-34 in both sexes. Distinct dynamics between men and women was evident in external deaths (Figure XX - XX). All age groups apart from 65-74 years in men exhibited significant seasonality across the study period, though the 55-64 age group seasonality only began to emerge in the early 1990s, with the 75-84 years group disappearing at 12 months by the end of the study period (Figure XX - XX). Women, in contrast, bore no significant 12-month seasonality from 45-64 years. While 25-34 year olds were significant for the entire period, the 12-month seasonality had disappeared by the mid 2000s (Figure XX - XX). 12-month significance in 65-74 years in women was also patchy throughout the time period.

3.4.2 Centre of gravity analysis

Death rates in men aged ≥ 45 years and women aged ≥ 35 years peaked in January and February, and were lowest in July and August (Figure XX). A similar temporal pattern was seen in children younger than five years of age, whose mortality was highest in February and lowest in August. In contrast, the peak and minimum of mortality in older boys and young men (ages 5-34 years) occurred in June/July and December/January, respectively.

For ages ≥ 45 , Cancer mortality for both men and women were maximum in December/January and were at a minimum in July (Figure XX). The rest of the age groups' uncertainties were too large

below 45 to draw any meaningful conclusion from them. Cardiopulmonary mortality consistently peaked in January/February and was at a minimum in July/August for both men and women, although uncertainty was too large in ages 5-24 to take the central figure as a representative statistic (Figure XX). External deaths in men were stable (i.e. with low uncertainty) from 0-64 and ≥ 75 years, although the 0-64 years deaths peaked in June/July and were minimal in December/January, with ≥ 75 demonstrating external deaths peaking in December/January and at a minimum for June/July (Figure XX). External deaths in women were stable from 0-34, where they peaked and troughed the same as men. Women ≥ 75 were the same as men also. Other deaths for men and women were stable at ≥ 35 years, at a maximum in December/January and a minimum in July/August (Figure XX).

3.4.3 Change in percentage difference in death rates between maximum and minimum mortality

Percent difference in death rates between mortality in peak and minimum months declined by less than seven percentage points for people older than 45 years of age from 1982 to 2013 (Figure XX). In contrast, the difference between peak (summer) and minimum (winter) death rates declined significantly in younger ages, by nearly 25 percentage points in males aged 5-14 years and 15-24 years. Under five years of age, percent seasonal difference declined by a statistically-significant 13.2 percentage points (95% CI 8.1 to 18.2) for boys but only a statistically insignificant 5.0 percentage points (-12.0 to 2.0) for girls.

Cancer possessed no significant changes in percentage difference for any age group and is not shown here. Cardiovascular deaths showed significant percentage increase 25-34 men, and a significant decrease in 0-4 boys (Figure XX). Significant decreases were evident in many of the age groups for both men and women in external deaths, though the 0-4 age group experienced an increase in both boys and girls (Figure XX). There was only one significant change in other deaths in 0-4 boys, and is not shown here.

3.4.4 Sub-national centre of gravity analysis

The sub-national centre of gravity analysis shows that mortality peaks and minima in different climate regions (Figure 1) are consistent with the national ones (Figure 5), indicating the seasonality is largely independent of geography. The relative homogeneity of the timing of maximum and mortality is evident, despite the large variation in temperatures that exist between climate regions during the same months. For example, in men and women aged 65-74 years, mortality peaked in February in the Northeast and Southeast, even though the average temperatures for those regions were different by over 13 degrees Celsius (9.3 in the Southeast compared with -3.8 in the Northeast). Furthermore, above 45 years of age, there is little inter-region variation in the percent seasonal difference, despite

the large variation in temperature difference between the peak and minimum months (Figure 6; none of the associations with temperature difference were statistically significant above 45 years of age). The observed geographical consistency in seasonal mortality variation in the USA, also seen in a study of 36 cities using deaths aggregated across age groups and over time,(16) contrasts from the pattern observed across Europe, where the difference between winter and summer mortality tends to be lower in the colder Nordic countries than in warmer southern European nations, possibly because the former have put in place environmental (e.g., housing insulation and heating) and health system measures to counter the effects of cold winters.(3, 4, 6) The absence of association between the magnitude of mortality seasonality and seasonal temperature difference indicates that different regions in the USA are similarly adapted to temperature seasonality.

3.5 Discussion

In a novel analysis of mortality seasonality, we used wavelet and centre of gravity analyses, which allowed not only systematically identifying and characterizing seasonality, but also examining how it changes or disappears over time. Analyzing seasonality over three decades in relation to age, sex, and geography allowed us to identify distinct seasonal behaviors in relation to age and sex, including the higher summer mortality in young men which has rarely been reported,(17) and to establish that mortality seasonality is consistent sub-nationally in terms of both timing and magnitude. Insights of this kind would not have been possible analyzing data averaged over time or fixed to pre-specified frequencies.

The substantial decline in seasonal mortality differences in adolescents and young adults are related to diminishing role of external causes of deaths (as discussed above), especially from road traffic crashes, which are more likely to occur in the summer months,(20, 21) and are more common in men. The weakening of seasonality in children under five years of age may be related to the reduction of deaths from respiratory causes, which have a strong seasonality, and the accompanying increase in the proportion of deaths that occur in the neonatal period, which do not vary noticeably throughout the year.(17, 20, 22, 23) Further elucidation on the causes of deseasonalization in these age groups requires analyzing changes in the composition of causes of deaths, as well as shifts in seasonality of causes of death themselves.

In contrast to young and middle ages, mortality in older ages, where death rates are highest, maintained persistent seasonality over a period of three decades (we note that although the percent seasonal difference in mortality has remained largely unchanged in these ages, the absolute difference in death rates between the peak and minimum months has declined because total mortality has a declining long-term trend). This finding demonstrates the need for environmental and health service

interventions targeted towards this group irrespective of geography and local climate. Examples of such interventions include enhancing the availability of both environmental and medical protective factors, such as better insulation of homes, winter heating provision and flu vaccinations, for the vulnerable older population.(24) Social interventions, including regular visits to the isolated elderly during peak mortality periods to ensure that they are optimally prepared for adverse conditions, and responsive and high-quality emergency care, are also important to protect this vulnerable group.(4, 24, 25) In many countries such services are increasingly under strain in an era of austerity. Emergent new technologies, such as always-connected hands-free communications devices with the outside world, in-house cameras, and personal sensors also provide an opportunity to enhance care for the older, more vulnerable groups in the population, especially in winter when the elderly have fewer social interactions.(26) Such interventions are important today, and will remain so as the population ages and climate change increases the within- and between-season weather variability.

The strengths of our study are its innovative methods of characterizing seasonality of mortality dynamically over space and time; using wavelet and centre of gravity analyses; using ERA-Interim data output to compare the association between temperature and seasonality of death.

4 Objective 2: Impact of large-scale ambient temperature changes on cause-specific mortality by age-sex group in the USA

4.1 Background

While previous work has focused on how mortality could be affected from hot and cold episodes in the days after a heat or cold event,(15) there is a lack of evidence to suggest how living in a generally-changed climate will affect mortality patterns, both by demographic and geography. Humans have evolved and adapted to a steady-state climate over the past few thousand years. The rapid large-scale climate change which occurs over the next century or so could prove too fast for humans to adequately prepare for, and so deviations from long-term mean temperatures could have deadly consequences for communities around the world. [ref]

Previous studies have explored the relationship between temperature and mortality, but there remains a large set of climate phenomena which are not captured by looking at only the mean in isolation.(16,17) There also is evidence to suggest that changes in the mean and in the heaviness of the tails of a temperature distribution will be significant to the risk profile of human health. The number of extreme heat and cold anomalous events (e.g. heat and cold waves) and variability of weather could also cause changes in risk, and so should also be included in an analysis of future risk to climate change (Table 1).

Our work thus attempts to address a broader research question, i.e. what the excess risk of mortality could be for different parts of society under future temperature distributions potentially influenced by climate change. We have developed a framework to capture several potential changes climate change could cause. Also of interest to those reporting metrics in the meteorological community will be information on which statistics will be most relevant to health, as clarifying and simplifying data sharing is a key goal of the Lancet Countdown.(11,18)

This framework will also help address the question of how overall year-total mortality will be affected due to a changing climate, as there is ongoing discussion as to whether climate change could bring benefits to the numbers of deaths in a year.(19,20)

4.2 Data

The data sources are the same in this study as described in subsection 3.2.

4.3 Methods

4.4 Generation of temperature statistics

We obtained data on temperature from ERA-Interim, which combines predictions from a physical model with ground-based and satellite measurements.⁽²³⁾ Other datasets were considered, but were not used as they did not assimilate ground data. [ref MERRA-2 and NARR] We used gridded four-times-daily estimates at a resolution of 80km to generate monthly population-weighted temperature statistics by state throughout the analysis period. The method of taking daily averages to represent a measure of average heat stress has precedent.⁽⁵⁾

We developed a repertoire of summary statistics for a month based on daily values (Table X), built to comprehensively reflect both how climate change could affect long-term patterns of health, as well as weather. Where we looked at heat or cold anomalies, we used the mean daily temperature. Mean daily temperature is a good summary of day and night temperature, which is significant to include as the lack of relief during the night time is known as a key risk factor in premature mortality.⁽⁴⁾ We generated 30-year average temperature values for each state-month combination, as well as 10th and 90th percentiles, using 1980-2009 as the reference period to measure against.

Where each statistic summarised the number of days or episodes in a month, a scaled value was calculated to account for varying length of months, by multiplying each months value by a factor that would make it equivalent to a 31-day month.

Table 2: Names of temperature statistics, their descriptions and why they are of interest. All centred statistics are based around a state-month long-term normal ($^{\circ}\text{C}$), calculated from 1980-2009

Name	Unit	Description	Why of interest?
Mean	$^{\circ}\text{C}$	Deviation from state-month long-term normal	How shift in centre of temperature distribution may affect human health
Standard deviation		Standard deviation of temperature for a state-month	How much variation around the mean of a distribution may affect human health
10th Percentile		Deviation of the 10th percentile temperature of a month from state-month long-term normal	How change in heaviness of tails of temperature distribution may affect human health
90th Percentile		Deviation of the 10th percentile temperature of a month from state-month long-term normal	
Relative Warm Anomaly	Number of episodes	Number of episodes of warm anomalies (more than 3 days in a row above the 90th percentile of state-month long-term normal from 1980-2009)	How changes in episodes of anomalous temperature may affect human health
Relative Cold Anomaly		Number of episodes of cold anomalies (more than 3 days in a row below the 10th percentile of state-month long-term normal from 1980-2009)	

4.4.1 Statistical analysis

Initial empirical analysis showed distinct behaviour between neighbouring age-sex groups for monthly mortality and susceptibility to changes in climate. For this reason, as well as computational limitations, we ran each age-sex group model separately.

For each age-sex group, we used a Bayesian spatiotemporal model that was formulated to incorporate features of deaths rates in relation to location of residence, month of the year, and climate statistic, over space and time. A Bayesian structure allows the full distribution of the target parameters to be inferred, allowing for a more natural expression of the probability of changing risk from a changing climate. The model specification is provided fully in the appendix.

For all-cause mortality, death rates vary with the month of the year, with rates highest for older age groups of both sexes in the winter months, and lowest in the summer months, with the reverse true for younger men (as demonstrated in section XX above). The rate of change of deaths rates is also different when comparing months in the year. Therefore, we allowed each month of the year to possess a different mortality level and trend. We used a random walk structure for both monthly intercepts and slopes, which is widely used to characterise smoothly varying associations, as evident in monthly variation.

Death rates are also variable by state, both in intercept and trend. We thus allowed death rates to vary by state. In addition, states closer in geography might be more similar than those further away. We employed the Besag, York, and Mollie spatial model, described in the appendix and elsewhere, to reflect this. This allows death rates of a state and their trends to be estimated based on their own data as well as using those of their neighbours, producing more stable estimates of death rates and trends. The level to which information is shared between neighbouring states depends on the uncertainty of death rates in a state and the empirical similarity of neighbouring states.

We included a temperature summary term to infer risk nationally for a single age-sex group (equation 1) The terms in equation 1 are described in detail in section XX in the appendix. In brief, $\sum_{i=1}^n \gamma_i VAR_i$ in equation 1 describes the climate variable part of the model where n ($n = 0, 1, 2, 3...$) climate statistics can be included.

$$\begin{aligned}
\log(\mu_{m,s,t}) = & \alpha_0 + \alpha_{M[m]} + \alpha_{S[s]} + \alpha_{X[m,s]} + \\
& (\beta_0 + \beta_{M[m]} + \beta_{S[s]} + \beta_{X[m,s]})t + \\
& \sum_{i=1}^n \gamma_i VAR_i + \pi_t + \epsilon_{m,s,t}.
\end{aligned} \tag{1}$$

Since time trends can also be non-linear, we modelled the time trends using the linear terms described above along with smoothly varying non-linear terms, specified with a random walk term over time, as well as an overdispersion term to capture any additional non-linearity.

4.5 Results

4.6 Discussion

5 Objective 3: Risk assessment of large-scale ambient temperature changes using climate projections and IMAGE

5.1 Background

5.2 Data

The data sources are the same in this study as described in subsection 3.2.

5.3 Methods

5.4 Results

5.5 Discussion

6 Future work