Anomalous temperature and seasonality of mortality in the United States

Robbie Michael Ignatius Parks

Department of Epidemiology and Biostatistics
School of Public Health
Faculty of Medicine
Imperial College London

Thesis submitted for fulfilment of the requirements for

Doctor of Philosophy (PhD)

July 2019

Abstract

Measuring and identifying drivers of mortality are key functions of public health research. Information on seasonal mortality enables prioritising and evaluating interventions during particular times of the year. It also provides the basis to study the association of death rates with seasonal temperature. My thesis aimed to characterise seasonality of death rates by cause of death, month, sex and age group, and to understand the role of temperature in the observed patterns. I generated death rates from geo-coded vital registration data on all 85.5 million deaths in the entire contiguous United States over a 37-year period (1980-2016). I grouped the underlying causes of death into mutually exclusive and collectively exhaustive broad causes (cardiorespiratory diseases, cancers, injuries, other) along with further sub-causes for cardiorespiratory diseases and injuries. I processed ERA-Interim reanalysis weather data to create population-weighted monthly temperature statistics. I applied wavelet, centre of gravity and circular statistics methods to analyse seasonal dynamics of mortality. I found that overall death rates for those 45 years and older were highest in the winter, mostly due to cardiorespiratory disease deaths. In contrast, death rates in adolescents and young adults peaked in the summer, mostly due to injury deaths. Seasonal differences in older age groups have changed little over time, whereas in young children they have largely disappeared. I then formulated a Bayesian spatio-temporal model to estimate how anomalous temperature – defined as temperature deviation compared to long-term norm for each state and month – affects deaths from different causes. A 1°C anomalously warm year in the contiguous United States would be associated with an estimated 941 (95% credible interval (CrI) 831, 1,053) additional injury deaths (0.5% of total injury deaths in 2016), concentrated in adolescent to middle-aged males. There would be an estimated 4,369 (4,024, 4,706) fewer cardiorespiratory disease deaths (0.4% of total cardiorespiratory disease deaths in 2016), concentrated in those 55 years and older. There would be a decrease of cardiorespiratory disease deaths in all but summer months. I found no association between anomalous monthly temperature and cancer deaths. Continued efforts are necessary to address seasonal peaks in mortality, especially in older age groups, across the United States, especially in a changing climate.

Table of Contents

List of tables

List of figures

List of equations

List of abbreviations

3DVAR	Three-dimensional variational data assimilation
4DVAR	Four-dimensional variational data assimilation
ACME	Automated Classification of Medical Entities
BUGS	Bayesian inference Using Gibbs Sampling
CAR	Conditional autoregressive
CDC	Centers for Disease Control and Prevention
CKDu	Chronic kidney disease of unknown aetiology
COPD	Chronic obstructive pulmonary disease
CI	Confidence interval
CrI	Credible interval
DHS	Department of Homeland Security
DIC	Deviation information criterion
ECMWF	European Centre for Medium-Range Weather Forecasts
EPA	Environmental Protection Agency
ERA	ECMWF re-analysis
GMRF	Gaussian Markov random field
ICD	International Classification of Diseases
IHD	Ischaemic heart disease
INLA	Integrated nested Laplace approximation
IPCC	Intergovernmental Panel on Climate Change
JAGS	Just another Gibbs sampler
MCMC	Markov Chain Monte Carlo
NCAR	National Center for Atmospheric Research
NCEP	National Centers for Environmental Prediction
NCHS	National Center for Health Statistics
NOAA	National Oceanic and Atmospheric Administration
NUTS	No U-Turn Sampler
$PM_{2.5}$	Particulate matter with diameter less than 2.5 micrometres
PM_{10}	Particulate matter with diameter less than 10 micrometres
PRISM	Parameter-elevation Regression on Independent Slopes Model
SDG	Sustainable Development Goal
TMB	Template Model Builder

UN	United Nations
UNFCCC	United Nations Framework Convention on Climate Change
UNISDR	United Nations International Strategy for Disaster Reduction
VR	Vital registration
WHO	World Health Organization
WMO	World Meteorological Organization

Copyright

Declaration of originality

I hereby declare that the work in this thesis is my own original research and that I have appropriately cited any work that is not my own.

Acknowledgments

First and foremost, I would like to thank my supervisors Majid Ezzati and Ralf Toumi. Majid, you have been a great source of guidance and inspiration throughout my time working with you. You have really helped me through some great challenges, intellectual and personal alike. I honestly do not know how I would have got through what life threw at me during these past few years without your support. For that I will always be extremely grateful. Ralf, thank you for the wisdom and enthusiasm you have shared when approaching my research questions. I always came away from our discussions feeling that I had learnt a great deal.

I would also especially like to thank my Bayesian spatio-temporal colleagues, James Bennett, Helen Tamura-Wicks and Vasilis Kontis. You have sat through countless hours of discussions about my work and your contributions were always essential. You each have the patience of a saint. In addition, my thesis was made possible by the invaluable advice and comments from my colleagues, especially Honor Bixby, Bin Zhou, Kyle Foreman, Sierra Clark, Cristina Taddei, Rodrigo Carillo Larco, Nia Iurilli, Niloofar Shoari, Bethlehem Solomon, Jonathan Pearson-Stuttard, Mariachiara di Cesare, James Bentham, Felipe Teixeira de Almeida, Nathan Sparks, Shuai Wang, Theo Rashid, Emily Muller, Ricky Nathvani, Andrea Rodriguez Martinez, Marisa Sophiea, Queenie Chan and Aubrianna Zhu. Thank you also to my long-suffering loved ones, especially Charlie Roscoe, Fr. Pat Browne, Rafael Gafoor, Simon Shillaker, Chris O' Shea, Mervyn Alaras and Glitches (Michael O' Neill and James Warner).

Christy you are the best sister I could ever hope for. Thank you for putting up with my idiosyncrasies these past few years. No more late-night concentration music emanating from my room after this I promise. Anna Rommel, Elfie and family: thank you for all the love and for believing in me. Finally, mum and dad, thank you for everything you gave me over the years. I know that, somewhere and somehow, you are both watching and willing me on.

1 Introduction

1.1 Rationale

Large-scale annual programmes in advanced temperate countries like the United States are concerned with minimising the peak in deaths observed every winter.[1] Measuring trends in seasonal mortality is therefore a key input to public health research. While mortality seasonality is well-established, there is limited information on how seasonality varies by cause of death, sex, age group, as well as over time and space. This information is needed to better identify and target at-risk groups during periods of elevated risk.

There are well-established links between daily temperature and mortality from several causes of death, particularly deaths from cardiorespiratory diseases.[2–7] Further analysis is needed to understand the impact of these temperature shifts across a full range of causes of death, with a study design which will enable identification of particular vulnerable groups.

Research in this area is also motivated by global climate change, which will cause deviations from long-term norm temperatures and have wide-ranging impacts on health.[8] The Paris Agreement of 2015, within the framework of the United Nations Framework Convention on Climate Change (UNFCCC), mandated that 'Parties should, when taking action to address climate change, respect, promote and consider their respective obligations on the right to health'.[9] Further, the Lancet Commission on health and climate change, established in 2015, made a primary recommendation of more investment in climate change and public health research.[10]

1.2 Aims and objectives

The aim of my thesis is to estimate trends in seasonal mortality and how shifts from long-term norm temperatures may impact mortality by cause of death, sex and age group in the United States. The research has two objectives that will help achieve this aim:

1) to identify and quantitatively describe seasonal dynamics of mortality; and
2) to develop a mathematical model to establish how monthly mortality is associated with deviations in temperature from long-term norms.

1.3 Structure of the thesis

In Chapter 2, I review the previous research related on temperature and mortality and methods related to spatial and temporal modelling. In Chapter 3, I summarise the temperature and mortality data for the United States and introduce the anomalous temperature metric. In Chapter 4, I present findings on the seasonal dynamics of mortality in the United States. In Chapter 5, I introduce the statistical model used to quantify the association of mortality with anomalous temperature. In Chapter 6, I present the findings of the association of injury mortality with anomalous temperature. In Chapter 7, I present the findings of the association of mortality from cardiorespiratory diseases and cancers with anomalous temperature. In Chapter 8, I will conclude with a discussion on how my results compare with existing literature, the implications of the findings from a public health perspective and potential future directions.

2 Background

2.1 Trends in mortality in the United States

Accurate measurement of mortality is an essential public health function. Reliable mortality data can influence policy decisions and inform whether countries are on track to achieve targets such as the Sustainable Development Goals (SDGs)[11] as well as national and subnational goals. However, generating reliable mortality data requires considerable effort in accurately recording the number of deaths by cause and measuring the size of a population. In England and Wales, calculating mortality statistics goes back several centuries.[12] In the United States, records span over a century.[13] However, in many other countries today, mortality data continue to be unreliable or unavailable.[14]

Over the 20th century, the burden of mortality in the United States has largely shifted from communicable towards non-communicable diseases, such as cardiorespiratory diseases and cancers.[15] Mortality from most causes of death in the United States is on a long-term downwards trajectory,[16] though recent trends show an increase in seven of the ten leading causes of death.[17] Death rates stratified by sex and age group are often a particular focus in efforts to quantify mortality, which can then be processed to generate life expectancy by applying life table methods.[18] Life expectancy in the United States, like other high-income countries,[19,20] has steadily increased over the last few decades,[16] though not all groups have shared this improvement,[21–23] and life expectancy in the United States has recently slightly decreased.[17] Improvements in income, nutrition, education, sanitation, health care and air quality have brought the United States to a life expectancy at birth of 76.1 years in men and 81.1 years in women in 2016.[24] Some predict that an overall long-term increase in life expectancy will continue, albeit at a slower pace,[25,26] while others contend that a further decrease is imminent due to health risks as well as climate change.[27–29]

2.2 Trends in temperature in the United States

In meteorological terms, recording weather (or short-term variation in meteorological conditions) is critical for understanding climate (or long-term meteorological conditions, usually an average of at least 30 years) in context.[30] The importance of recording and understanding weather is long-recognised. The ancient Sanskrit texts in the Upanishads from 3000 BC contain theories on how seasonal cycles in temperature may be caused by the Earth's movement around the sun.[31] However, the invention of the liquid-in-glass thermometer and barometer in the 17th century provided the foundation for modern collection of temperature data.[32,33] In some places, such as throughout the former British Empire, weather data have been collected from weather stations for over a century, with the British Isles' records dating back to the 1650s.[34] Temperature records in the United States go back to 1895 in some areas.[35] Land-based weather stations in the United States are now widespread.[36]

The United States is a large temperate country possessing distinct climate regions,[37] with annual average temperatures within the climate time-scale of 1980 to 2009 ranging from 7.6°C in the West North Central region, to 18.4°C in the Southeast.[38] Annual average temperatures in the United States have increased by 0.7-1.0°C since records began in 1895.[39–43] The greatest and most pervasive warming has occurred in winter months, with average increases of over 0.8°C.[35] Summer months experienced less consistent changes, with warming mainly along the western parts of the country, and some cooling in the South and Midwest.[35] Under future climate change, annual average temperature in the United States is projected to increase by 1.4°C by the middle of the 21st century (relative to the late 20th century reference period), with up to a 6.6°C increase by 2100.[44] All regions of the United States are projected to experience significant warming, with the greatest warming expected in the northern half of the country.[44]

Heat waves are multi-day extreme heat events which have no standard definition, but typically take place over consecutive days over a chosen temperature threshold.[45] However they are reasonably defined,[45] heat waves have increased in frequency and length over the past century throughout the United States.[46,47] One study with a dataset covering the entire United States from 1961 to 2010 found that the number of heat waves annually in American cities has increased by 0.6 heat waves per decade; the length of heat waves has increased, and the length of the heat wave season has increased by six days per decade.[46] Exposure to extreme heat events in the United States is projected to increase four- to six-fold by the end of the 21st century compared to the end of the 20th century.[48] This increase in dangerous heat wave exposure is similarly projected worldwide.[49]

2.3 Seasonal mortality

Death rates vary by season, and on monthly, weekly and daily scales.[50] Seasonal variation of death rates within a year has been studied since ancient times in the West. The first record of this was when the Greek physician Hippocrates wrote about the impact seasons have on deaths.[51] A modern study has also shown that deaths were seasonal in ancient Rome.[52]

2.3.1 Excess winter mortality

It has been conclusively established that in temperate countries, such as in Europe and the United States, death counts from all causes of death are higher overall in the winter than in the summer.[53–55] This phenomenon is referred to as excess winter mortality or excess winter deaths. Excess winter mortality is defined either as the ratio or the absolute difference between the number of deaths in winter months (usually December to March) and those in the rest of the year. Excess winter mortality has become a common metric to quantify seasonal differences in mortality, exemplified by its use in annual statistical bulletins by the Office for National

Statistics in the United Kingdom to give a measure of the severity of a year's winter on mortality,[56] as well as in similar research in the United States.[54] Excess winter mortality for deaths from all causes in temperate high-income countries in Europe ranges from 5% to 30%.[53] In the United States, excess winter mortality is up to 40% in older age groups.[38] An assumption of excess winter mortality is that the number of deaths is highest in the winter. This is invalid when examining seasonal dynamics of mortality in certain causes of death such as unintentional injuries and may also be inaccurate for certain sex, age group and location combinations within a cause of death.[57] The usefulness of excess winter mortality to measure the impact of cold weather has also been questioned, as the metric assumes that all cold-attributable deaths occur in the winter months.[58,59]

2.3.2 Methods of analysing seasonal mortality

Alternative ways of assessing seasonal mortality exist. Studying the United States, some have employed algebraic techniques to examine seasonal mortality stratified by demographic group, cause of death and metropolitan areas.[50,60–63] Others have used Fourier spectral analysis or cosinor functions to fit seasonal models to monthly death rates in several countries.[64–68] One limitation of these analyses is the assumption that seasonal death rates have a fixed 12-month cycle. One study used a more flexible approach by including certain other pre-determined cycle lengths in an assessment of seasonality of mortality in Japan.[69] Another disadvantage of these types of analyses of seasonal mortality is that the difference between levels of maximum and minimum mortality in a year is assumed to be constant over the study period. However, over time, behaviour of seasonal mortality can change or disappear entirely. Lexis diagrams, which are two-dimensional plots of age against time used to show births and deaths that occur over time, have been adapted to analyse changes in seasonality of mortality over time.[70]

2.3.3 Determinants of seasonal mortality

The existence of seasonal variation in mortality has stimulated research into why death rates vary over a year. Understanding the dynamics and drivers of seasonal mortality is required for planning interventions, which would aim to reduce the severity of peaks in death rates. Incidence of influenza, especially in the elderly, is considered an important determinant of winter mortality in temperate countries such as the United States.[71,72] Influenza incidence in the United States is also a good predictor of cardiovascular disease mortality, which also peaks in winter months.[73,74] However, targeting of the elderly with influenza vaccinations in high-income countries has shown limited success on influencing seasonal mortality.[75,76] This may be due to their underutilisation, with 37.1% of adults covered in the 2017-2018 season in the United States.[1] The vaccine effectiveness was 38% for this period.[77] Some have suggested that improvements in current vaccine production may yet further reduce influenza mortality in winter months.[78,79] Some influenza vaccines are cultured using chicken eggs, which is posited as leading to mutations in the vaccine candidate and lower effectiveness as a vaccine in the general population.[78] An interaction between temperature and influenza in influencing seasonal increases in death rates in winter is also hypothesised.[80]

Since temperature varies in a similar way to deaths from some causes throughout a year in temperate countries like the United States, some have investigated the link between them. If temperature were a primary driver of the seasonal mortality, places with larger average temperature ranges throughout a year may show larger excess winter mortality. This was not seen in a study of 36 cities in the United States.[81] A similar study of European countries found that the difference between winter and summer mortality was lower in the colder Nordic countries than in warmer southern ones.[53] Others argue that weather is the main driver of winter deaths and that previous calculations using excess winter deaths is confounded by not excluding cold-attributable deaths outside the winter period or accounting for inter-annual variability of winter length.[58,59] Beyond temperature variation and influenza severity, a further proposal is that additional drivers of seasonal mortality may be due to a more advanced chain of causality. This would involve social factors (such as housing conditions), biomedical reaction to climate (such as infections of respiratory systems) and demographic reaction to biomedical changes (such as changed mortality risk in susceptible groups).[57]

Beyond a seasonal cycle of mortality, inter-annual variation in mortality exists. For example, death rates in one or more months in a particular year may experience a deviation from the long-term smooth seasonal cycle. Temperature is well-established as a major source of variation in mortality and morbidity.[3–7]

2.4 Temperature and mortality

2.4.1 Effects of temperature on humans

Humans, like other successful species on Earth, have 'inherited or acquired the behavioural, morphological, and physiological attributes necessary to avoid, tolerate, and adapt to the stresses of life'.[82] This process has included adapting to local climate along with its seasonal cycle. The Intergovernmental Panel on Climate Change (IPCC) defines climate adaptation as 'the process of adjustment to actual or expected climate and its effects'.[83] Despite humans' successful adaptation, the potential for unusual or extreme weather to disrupt society and health has been demonstrated and recognised throughout history. Exceptional temperature events are known via numerous European town chronicles in the sixteenth century.[84] Measurements in Europe of the past 500 years also demonstrate further evidence of temperature anomalies.[85] More recent periods of extreme heat, such as the 1995 Chicago heat wave,[86,87] or 2003 European heat wave,[88] claimed many victims.

The human body is in a state of normothermia, or comfortable resting temperature, between 36.5-37.5°C.[89] Heat stress, or the 'perceived discomfort and physiological strain associated with exposure to a hot environment', occurs at temperatures above this range.[90] Healthy adults have efficient corrective mechanisms to regulate high body temperature by vasodilation and perspiration.[6] However, even a healthy human body has an upper limit to its endurance of excessively warm temperatures.[89]

2.4.2 Extreme heat and cold

When heat stress becomes extreme, this can result in potentially deadly medical conditions such as heat stroke, heat exhaustion, heat syncope, heat rash and heat cramps.[91] In the most acute cases, heat stress can lead to multiple organ failure and rapid death.[6] The impact of periods of extreme heat on mortality is substantial. However, direct causes of death from extreme heat stress only make up a small proportion of deaths attributable to daily elevated temperatures.[2] Most temperature-related daily deaths are instead attributable to non-extreme deviations from acclimatised temperatures, which suggests other direct and indirect pathways to mortality from elevated temperatures exist.[2–4]

Though extreme heat was not included as a natural disaster under the United Nations (UN) Sendai Framework for Disaster Risk Reduction of 2015,[92] there is a growing public awareness that prolonged periods of heat are dangerous and their worst effects need to be curbed.[6] Projections of the future climate predict an ever-greater challenge of handling elevated heat exposure worldwide.[49] In the United States, extreme heat is on average estimated to cause more deaths than any kind of weather-related hazard.[93,94] This is true even of a year like 2018,[95] when flooding in Maryland, hurricanes Florence and Michael and the California wildfires occurred.[96] Records from the United States Environmental Protection Agency (EPA) claim that from 1979, over 9,000 deaths have been attributable to extreme heat.[97]

There has been interest in establishing whether heat waves act as a modifier to the deadly effect of heat stress.[98–101] Heat waves can be described in different ways. The minimum threshold temperature of a heat wave can be defined either relatively (e.g., certain number of days above 95th percentile of local long-run temperatures) or absolutely (e.g., number of days above 30°C).[45] The minimum number of consecutive days required above a heat wave threshold to qualify as a heat wave also varies, with studies using two days or more.[45] The heat stress metric can also vary, with apparent temperature or various heat indices previously used instead of temperature.[45] The temperature threshold can also be defined by daily mean, minimum, maximum temperature, or temperature at a particular time of day.[45] Some studies have found an added effect for consecutive extreme heat days beyond the contribution of individual days.[98,100,102] Another study found that the total risk of consecutive days of elevated temperature can largely be explained by aggregating the independent effects of individual days' temperatures.[99] Others have found that estimates of the impact of a heat wave on mortality are sensitive to the definition of a heat wave itself.[45]

Prolonged periods of cold weather, which along with extreme heat is not included in the UN Sendai Framework,[92] are also a deadly risk to human health.[103] Below the range of normothermia, cold weather puts strain on the heart and respiratory system.[89] Extreme cold stress beyond the human body's ability to adapt leads to hypothermia and can quickly lead to complete organ failure and death.[89] Mild cold, however, leads to more cold-related deaths than extreme levels of cold.[2] In the United States, there is evidence that the severity of winter storms have decreased in the South and lower Midwest but have increased in the Northeast and upper Midwest.[104] While cold weather is also known to have a large impact on mortality,[2,103] global climate change will result in stronger and, on average, warmer deviations from long-term norm temperatures.[105]

A study of cold waves in the United States, analogous to heat waves but for extreme cold, found no added effect beyond the contribution of individual days' risk.[102]

2.4.3 Shape of temperature-mortality relationship

The shape of the temperature effect for daily mortality is non-linear and is variously described as a J-, U-, or V-shaped curve, with elevated risk both for warmer- and colder-than-average conditons.[2] Several designs of fitting temperature-mortality relationships exist with linear, polynomial and cubic spline curves.[106] With daily mortality, however, a death attributable to temperature on any given day may be due to the accumulation of exposure from a number of preceding days, of which the heat effect is more immediate than the cold effect.[2] Distributed lag non-linear models were developed to account for the non-linear and delayed effects in temperature and daily time-series data.[107] The relationship between temperature and mortality on a monthly scale is less extensively studied. There are a few examples of studies using monthly temperatures and mortality using a linear or log-linear relationship.[3,5,108–110]

2.4.4 Cause of death

Temperature variation has been associated with increased risk for all-cause mortality and mortality excluding injuries,[2,3,5–7,87,88,98–101,103,109,111–119] as well as for cardiorespiratory diseases.[3,5–7,87,88,115–117,120–127] In particular, studies have found an association between temperature and deaths from cardiovascular diseases,[7,87,115,117,123] including ischaemic heart disease (IHD),[103,117] myocardial infarction,[117,128] cerebrovascular diseases[7,103,117] and heart failure.[117] Causality behind these associations has been detailed in previous work.[103,129–131] In brief, higher-than-average temperature in warm months puts strain on the circulatory system by reduction of plasma volume from the release of platelets into the blood stream, as well as water and salt loss from sweating, which can cause artery blockages and sudden decreases in blood pressure.[131] Previous studies have shown that warmer days in summer months increase cardiovascular deaths, while not increasing hospital admissions.[130] This suggests that those who died in warmer weather may have already had pre-existing conditions, which made them more vulnerable.[130]

In colder conditions, warmer-than-average temperatures decrease cardiovascular deaths. This has been attributed to the inverse relationship between temperature and blood pressure.[129] Blood pressure increases during colder weather; higher blood pressure alters the ratio between supply and demand of oxygen delivered to the myocardium in the heart, which leads to greater stress in the ventricular wall and increases the work the heart needs to do.[129] Higher blood pressure also reduces mechanical efficiency of the heart and can impair blood flow in it, which may lead to myocardial ischaemia.[129] Vasoconstriction can also affect the ratio between systolic and diastolic blood pressure, producing vessel wall deformation and damage.[129] Other factors such as increased blood clotting and thrombosis may also have an influence on the increasing risk of cardiovascular disease death during colder weather.[103,129]

Increases in deaths from respiratory diseases are also observed in warmer than comfortable temperatures,[7,103,115,117,123] including chronic obstructive pulmonary disease (COPD),[117] asthma[117] and respiratory infections.[7,117] These are believed to occur in warm weather due to airway inflammation, and with elderly patients with existing COPD unable to dissipate excess heat.[130] The association of cold temperatures with respiratory diseases may be accounted for by a suppression of immune responses by stress hormones during cold weather, which reduces the body's resistance to infection.[103]

A few studies relating injuries to temperature exist.[108,117,132,133] Removing injury deaths from a dataset has been a common part of data pre-processing for temperature-mortality studies.[2,3,5,98,99,101,114,116] High temperatures and increased work-related injuries have previously been examined, with manual workers such as farmers, construction workers, firefighters, miners, soldiers and those in manufacturing roles highlighted as at increased risk.[134] A previous study of the United States has linked rising temperatures with feeling of despair, potentially leading to more suicides.[108]

Some studies have related temperature to some other causes of death.[7,108,116,117,126,128,135–139] Deaths from endocrine diseases,[117] diabetes,[116,117] genitourinary diseases[7,117] and some psychiatric disorders[108,116,117] have been identified as being sensitive to temperature. Deaths from cancers have been studied in association with temperature, with largely no relationship observed.[126,136,139] An older study which observed an association between cancer deaths and temperatures urged caution in interpretation of the results, as the study did not take into account local long-term temperature norm differences.[136] Recent worldwide increases in chronic kidney disease of unknown origin (CKDu), which have been observed in agricultural communities in particular,[140] have also been associated with rising temperature,[135] along with other possible causes such as increasing use of pesticides.[141] Others argue that there is not enough evidence to draw any firm conclusions about the origin of increases in CKDu incidence.[140,141]

2.4.5 Demographic determinants

Men and women have different physiological responses to heat stress.[142] Some studies have analysed separately by sex to establish if the temperature-mortality relationship is disintct.[3,5–7,88,100,108,109,117,121,122,127,143–149] Women have been found to be more vulnerable than men from daily increases in temperature for all-cause mortality and cardiorespiratory disease deaths,[3,5,88,100,116,122,127,143,147–149] whereas men were found to be more vulnerable for circulatory causes in one study.[145] Other work which explored differences between men and women's temperature-mortality relationship found no differences in vulnerability between men and women,[108,109,121,144,146] though analysing deaths from all ages together. A recent review recommended further research by sex due to lack of existing evidence in differentials of the temperature-mortality relationship.[7]

The association between temperature and mortality also varies by age group. Where study design enabled comparison, greater vulnerability to rising temperatures was found for all-cause and cardiorespiratory disease mortality in older age groups and the elderly compared with healthy adults.[3,5,6,115,117,120–122,127,128,143,144,147,149–152] Children and adolescents were also shown to be at greater risk than healthy adults.[3,5,121,145,153,154] Further work on differentials between age groups in temperature-mortality relationship is also recommended in a recent review.[7]

Urban-rural differences have been found in some studies, with those in urban areas more at risk of temperature increase.[127,143] Other studies have also examined the differential impact of temperature by race and found that non-whites were at greater risk than whites in the United States,[121,155] though some have not found any meaningful difference.[109,123] Some socioeconomic factors were found to potentially elevate risk, such as living in a lower income area, having a lower education, and increased poverty,[3,156,157] though some studies found no association.[123,143]

2.4.6 Other environmental determinants

Aside from air temperature, various other heat stress indices have been developed. These include apparent temperature, humidex, heat index, dew point temperature among others.[158] These metrics use meteorological variables besides temperature, such as humidity and wind.

The main stimulus for suggesting these alternatives was that air temperature alone may not be the best predictor or a human's skin temperature, which plays a major role in a body's temperature regulation.[159] However, a study which attempted to systematically examine each of the proposed alternative heat stress metrics failed to find any that improved upon using the air temperature alone to predict mortality.[158]

Other environmental measures, such as air pollution, including ozone, particulate matter ($PM_{2.5}$, PM_{10}), carbon dioxide, sulphur dioxide and nitrogen dioxide, have been included in some studies as covariates to establish whether they may act as confounders to the relationship between mortality and temperature. A review of these studies found that no significant confounding or effect modification was evident on the inclusion of pollutants for the majority of the studies.[3] It has also been argued that pollution should not be a confounder of temperature-mortality association, as adjusting for pollution excludes the interaction between temperature and pollution.[160]

2.4.7 Mortality displacement

When a spike in deaths occurs, but deaths in the spike are only brought forward by a short time, the number of deaths after the spike will be lower than average. This phenomenon is known as mortality displacement.[3] Some speculate that this occurs when increased mortality occurs during elevated temperatures. The implication is that the increase in deaths would not necessarily be of public health significance, since the deaths were only slightly brought forward in time. While some studies claim to have found mortality displacement in temperature studies,[123,161,162] others claim to have found none.[108,163] One publication[163] directly criticised another study's methods.[162] A study of suicides found no mortality displacement when associating mortality with temperature.[108]

2.5 Modelling methodology and computation

2.5.1 Treatment of time in models

To establish a dataset for use in a study, often data across several years are collected. A large body of scientific work focuses on the association of daily mortality and morbidity with episodes of extreme heat lasting up to a few days.[3–7] Various methods have been used for the time series data. The time series regression model uses techniques originally developed in pollution modelling to attribute mortality counts on a particular day to observed temperature and other time-varying factors, most commonly assuming an over-dispersed Poisson regression model.[106,164,165] Other studies use case-crossover design, where temperatures on the day of death and those immediately preceding are compared to control days where death did not occur.[166] Case-crossover designs are commonly used for analysing effects of short-term exposures, such as temperature.[166]

Some studies have associated monthly or seasonal average temperatures with mortality.[108,109] Other studies have accounted for seasonal impacts by stratifying daily data by month,[81] or allowing size of association with temperature to change throughout a season.[114] Kinney highlighted how surprising it is that analysing by month or season to avoid potential seasonal confounding has to date ‘rarely been applied in temperature-mortality literature’.[81] There is also evidence that the temperature-mortality relationship changes throughout a single season, indicating that monthly analysis of the temperature-mortality relationship is sensible.[114]

2.5.2 Geospatial modelling

Mortality records are often geo-coded in addition to recording the time of death. Spatial studies have analysed the effects of countless factors on mortality along with disease incidence and prevalence.[167] Much previous work with the temperature-mortality relationship do not account for the spatial nature of the mortality data available. In the United States, study areas have ranged from a single city or a few counties,[86,87,121,138,168] to a region,[109,169] to many sites across the country.[81,101,114,118,120,123,126,137,144,146,155,161,170] A recent study used the entire United States dataset of suicide deaths.[108] Studies across different countries also compare temperature-mortality response using a consistent method.[2,101,114] Studies which are spread across multiple sites have typically modelled each location, often cities, separately.[81,101,114,119,120,123,137,144,146,155,161,170]

An alternative to analysing specific locations independently would be to account for the proximity or contiguity of the spatial units. A Bayesian formulation of a spatial model allows the 'borrowing of strength' across spatial and other units. This can be useful to create more stable estimates of risk where certain spatial units may have a relatively small amount of deaths and population levels but are next to a similar, but higher population, unit. Some recent work has used a Bayesian spatial framework to examine vulnerability to temperature.[127,170,171] Other work with temperature and mortality has instead involved Bayesian random-effect meta-analysis to pool relative risks from different locations.[172,173]

2.5.3 Implementation of Bayesian models

Advances in computing over the past three decades have enabled Bayesian statistical inference to become more commonplace in medical and public health research. In the 1990s, Bayesian inference Using Gibbs Sampling (BUGS) and later WinBUGS employ a Gibbs sampling algorithm to perform Markov Chain Monte Carlo (MCMC) iterations.[174] This was followed by Just Another Gibbs Sampler (JAGS).[175] Stan, an alternative program, uses the No U-Turn Sampler (NUTS) algorithm to improve speed and efficiency of inference.[176] Integrated nested Laplace approximation (INLA), using the R-INLA software, offers orders of magnitude of computational efficiency improvement in Bayesian inference compared to traditional MCMC for latent Gaussian models.[177] Template Model Builder (TMB) is another program which also uses Laplace approximation.[178]

2.6 Summary

Death rates in the United States vary by, sex, age group and across time and space. Temperatures in the United States have steadily increased, with marked changes since the 1970s. Death rates often vary seasonally. A framework is necessary to understand how seasonality has changed over time, and how temperature is associated with death rates by cause of death, sex and age group in the United States. Previous work has established that temperature and mortality are associated, with most studies examining different aspects of the daily temperature-mortality relationship in the United States and elsewhere. Studies of the temperature-mortality relationship which have used data across multiple sites have mostly treated each analysis independently from the others.

3 Data sources

3.1 Overview

I needed to identify and process appropriate data sources of mortality and temperature in the United States. This would enable me to carry out my analyses on the dynamics of seasonal mortality in Chapter 4 and the association of mortality with anomalous temperature in Chapters 6 and 7.

In this chapter, I give an overview of monthly mortality and temperature data sources, both large datasets which required considerable effort to process. I also present a generalisable algorithm for finding population-weighted monthly meteorological statistics, converting grid square temperature into state-level temperature values. The temperature output from this algorithm formed an essential data component of my thesis. My method is capable of turning commonly-found large datasets of gridded meteorological variables, such as ERA-Interim, into district-specific summaries anywhere in the world with an available shape file.

3.2 Cause-specific deaths and population

First, I describe how I processed vital registration death records; how I made choices in categorising death records by sex, age group, over time and space; and how I grouped categories into a mutually exclusive and collectively exhaustive combination of causes of death.

3.2.1 Objectives

My objectives in designing the cause of death classification included capturing significant causes of death within the country; collecting mutually exclusive and collectively exhaustive causes of death; grouping causes that possess comparable seasonal behaviour; generating monthly cause of death rates by state, sex, and age group; and avoiding small numbers issues.

3.2.2 Sources of mortality data

Sources of data on causes of death are available in various forms. These include vital registration (VR) systems, population censuses, mortality surveillance systems, verbal autopsy studies, hospital data, police records, mortuary records, and household surveys.[179] These sources, however, are not consistently matched in terms of coverage (percentage of the population which is monitored), quality (accuracy of diagnosis in records) and completeness.[179] VR systems are recognised as the highest quality mortality record data source,[14] and VR systems with medically-certificated causes of death are optimal.[180] The United States possesses medically-certificated VR records extending several decades. This is in contrast to many other parts of the world, where an estimated 53% of deaths went unregistered in 2016.[11] The World Health Organization (WHO) currently recognises 68 countries that produce and share high-quality VR data.[14]

3.2.3 Population data

To create death rates, population data is also necessary, with the number of deaths in a particular spatial (e.g., country or state) and temporal (e.g., year or month) unit of analysis in the numerator and the respective population in the denominator (Equation 1):

$$death\ rate = \frac{deaths}{population}$$

Equation 1. Death rate calculation formula.

A census, whereby each member of a population is counted manually by registering members of a household, provides the most reliable way of obtaining population data. In the United States, the census is every ten years, with the most recent in 2010.[181] The next United States census is due in 2020. Plans for the 2020 census, however, are controversial due to the proposed citizenship question potentially discouraging immigrants from contributing.[182] Some analysts fear that this may lead to undercounting of over 5%.[183] While a census takes place only every ten years, methods have been developed which reliably interpolate population by county, sex and age group in intercensal years with census years as boundary conditions.[184] Population data in the United States is therefore publicly available by county, year, centred around July, from the National Center for Health Statistics (NCHS) bridged-race dataset for 1990 to 2016[185] and from the United States Census Bureau prior to 1990.[186]

In a country like the United States, death rates in humans vary throughout the year.[53,187,188] Publicly-available records of death from VR systems in the United States contain the month and day of week of death. It is therefore not possible to identify a death record by exact day and date. Nevertheless, a month timescale allows perceptible variations in sub-annual deaths.[57] When calculating monthly death rates, monthly population data counts are required, which are not commonly available. Generating monthly population requires an algorithm to interpolate from yearly values.

For age-standardised death rate calculations in Chapters 3, 6 and 7, I used the WHO world standard population to create weighted averages of death rates from age-specific death rates.[189]

3.2.4 Assigning causes of death

With VR data, each death is assigned a single underlying cause. This will typically take place shortly after a patient has died, after which a physician fills out the death certificate. In the United States, this will include filling out five or more lines listing the chain of events which led up to a death.[190] Then the final underlying cause of death is processed using an algorithm, such as the Automated Classification of Medical Entities (ACME).[191,192]

This underlying cause of death is coded using the international classification of diseases (ICD) convention. The ICD system is currently on its 11th revision (ICD-11), released by the WHO in 2018.[193] However, each country decides when to adopt the latest ICD codes for recording underlying causes of death. The United States is currently still using its own variant of the ICD-10 coding,[194] ICD-10 Clinical Modification (ICD-10-CM).[195] ICD-10-CM was adopted in the United States in 1999, nine years after being endorsed by the Forty-third World Health Assembly.[195] The ICD-10 revision of the ICD coding system contains thousands of five-digit codes to which a death can be assigned. The ICD-10 codes have a high level of detail. Using each code individually for an epidemiological study would quickly present issues regarding small absolute numbers of cases. Research in public health policy therefore often requires that these causes of death are grouped into larger families. Though the ICD revisions provide a methodology to convert ICD codes into chapters,[196] these groupings are too broad for specific public health interest.

The version of ICD coding used by physicians also varies every time a new coding system is updated, which can cause some issues of comparability across ICD revisions.[197,198] When research spans multiple decades of mortality records in one country, there will almost certainly have been a change in ICD coding convention during that time. This can cause unusual changes in trajectory of death rates for certain causes of death.[199] This is complicated by the increase in number of ICD codes, which has expanded from 200 original codes to nearly 15,000.[200] This requires creating a unifying list of causes that can be applied across ICD revisions using a bridge between ICD revisions.[197] Misclassification of cause of death within and across ICD coding revisions can also occur due to certification issues, with evidence that certain causes of death, such as HIV or diabetes, have been at times systematically underrepresented.[191,201–203]

I therefore needed to create a mutually exclusive and collectively exhaustive set of causes of death within an ICD revision, consistent cause of death mappings across ICD revisions, and coherent groupings of causes of death with similar seasonality.

3.2.5 Classifying causes of death

I used data on all 85,453,845 deaths in the contiguous United States (i.e., excluding Alaska, Hawaii and all other American off-shore insular areas) from 1980 to 2016. I chose this range because ICD codes changed in 1979 and 2016 was the last year of available data, as well as because the ERA-Interim dataset only begins in 1979. During this period, there were 400,331 deaths in Alaska and Hawaii, 0.5% of total deaths in the United States during this period. All data quoted on population and death from here on in will refer to the contiguous United States. I used data on deaths by sex, age, underlying cause of death and state of residence from 1980 to 2016 through the NCHS (https://www.cdc.gov/nchs/nvss/dvs_data_release.htm). I used data on population from the NCHS bridged-race dataset for 1990 to 2016,[185] and from the United States Census Bureau prior to 1990.[186]

I grouped the underlying cause of death according to the ICD system (in the United States, 9th revision of ICD from 1980 to 1998 and 10th revision of ICD thereafter). Based on the Global Burden of Disease 2016 cause of death structure,[204] I developed mappings of ICD-9 and ICD-10 to a mutually exclusive and collectively exhaustive set of four broad causes and several sub-causes of death which are of public health interest in the United States and elsewhere. Grouping the deaths in this way provides a relatively detailed view of cause composition while avoiding small number concerns in modelling given the monthly and subnational nature of the data.

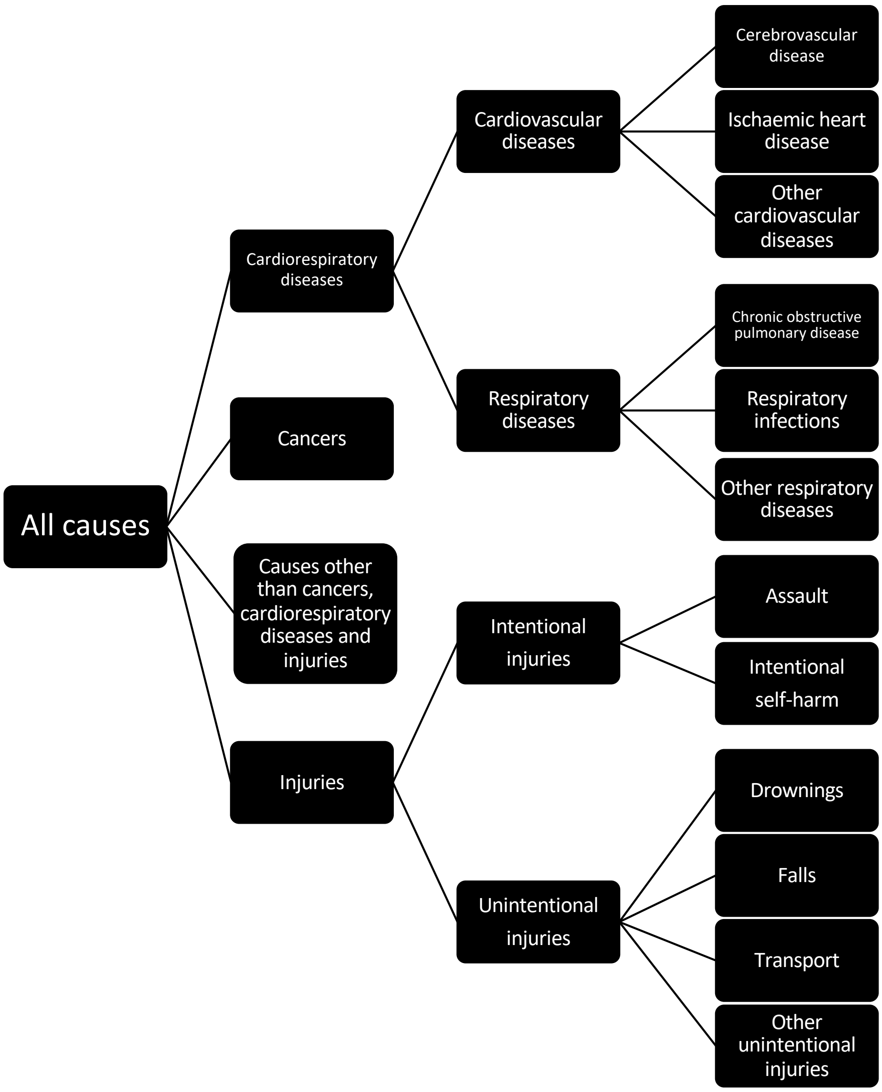

Figure 1. Tree structure of causes of death.

I adapted the categories to those which would exhibit common monthly variation and with outcomes relevant to changes in temperature, as well as being relevant to the United States burden of disease. For example, while deaths from cancers are a significant burden of disease in the United States and globally, with an estimated 8.9 million deaths worldwide in 2016,[204] death outcomes are not generally considered as sensitive to temperature changes, and so are all grouped as 'cancers' here. In comparison, cardiorespiratory diseases are also currently the leading cause of death in the United States, with cardiovascular diseases alone responsible for 900,000 deaths in 2016.[204] Sub-causes of cardiorespiratory diseases also exhibit different seasonal properties by sex and age group,[38] and therefore deaths in these sub-causes may also have distinct associations with temperature. Injuries and injury sub-causes are of interest, as hitherto fewer studies have been made for these causes than others, even though death rates from injuries vary seasonally,[38,57] and so there may be an association of their death rates with temperature change.

I show the tree structure in my mutually exclusive and collectively exhaustive cause of death hierarchy in Figure 1. The detailed causes of death can be aggregated to higher levels, such as grouping together all cardiovascular diseases or types of injuries. The causes of death can also be further aggregated up one additional level, such as grouping all cardiorespiratory diseases.

I used several competing models to generate monthly population estimates using yearly values, one of which modelled linear growth in months between years, with other more complex models such as exponential growth. The monthly population results were not sensitive to my choice. I calculated monthly population counts through linear interpolation, assigning each yearly count to July.

3.2.6 Age groups

I divided the data by sex and the following ten age groups: 0-4, 5-14, 25-34, 35-44, 45-54, 55-64, 65-74, 75-84, 85+ years. I did this in part to avoid small numbers, as well as to minimise the impact of potential age misclassification.[205] I calculated monthly death rates for each sex and age group, both nationally and for the 48 contiguous mainland states and District of Columbia. Death rate calculations accounted for varying length of months, by multiplying each month's death count by a factor that would make it equivalent to a 31-day month. This factor multiplication process was reversed for any additional death estimation from anomalous temperature rise in Chapters 6 and 7.

3.2.7 Patterns and trend in mortality

In the period 1980-2016, 20,070,797 boys and men and 21,034,212 girls and women died from cardiorespiratory diseases in the contiguous United States, accounting for 46.3% and 49.9% of all male and female deaths respectively (Figure 2). Cardiorespiratory diseases accounted for 54.3% and 58.9% of total deaths in 1980, and 39.9% and 40.8% of total deaths in 2016, for males and females respectively. Seasonality is apparent for total cardiorespiratory disease deaths, with a peak in the winter (Figure 3). I will further explore the presence of seasonality by sex and age group for various causes of death in Chapter 4.

Over time, cardiorespiratory disease death rates have decreased overall (Figure 4). Age-standardised death rates have decreased dramatically by over 50% during this period. This is replicated by most major sub-causes of death within cardiorespiratory disease deaths (Figure

48 in Chapter 7). Deaths from cardiovascular and respiratory diseases have previously been associated with variations in daily temperature from long-term norms.[2–7,127]

10,428,202 boys and men and 9,434,558 girls and women died of a cancer in the contiguous United States, accounting for 24.0% and 22.4% of all male and female deaths respectively (Figure 2). Deaths from cancers accounted for 21.3% and 21.2% of total deaths in 1980, and 23.1% and 21.6% of total deaths in 2016, for males and females respectively, peaking in proportion of total deaths for males in the 1990s and 2000s then returning back to a lower value. There were slightly more deaths from cancers in the winter than in the summer (Figure 3). Over time, death rates from cancers have decreased overall (Figure 4).

4,006,454 boys and men and 1,757,862 girls and women died from an injury, accounting for 9.2% and 4.2% of all male and female deaths respectively (Figure 2). Deaths from injuries accounted for 10.7% and 4.7% of total deaths in 1980, and 9.6% and 4.9% of total deaths in 2016, for males and females respectively. Seasonality in total deaths was apparent for injury deaths, with a peak in the summer (Figure 3).

The remainder of the deaths were from other causes. These consisted of a heterogeneous group of causes of death, within which the cause that led to death varied by sex and age group. Causes of death in this group included infectious and parasitic diseases, endocrine diseases, nutritional and metabolic diseases, immunity disorders, mental disorders, diseases of the nervous system, diseases of the digestive system, diseases of the genitourinary system, complications with pregnancy, skin diseases, diseases of the musculoskeletal system and connective tissue, and congenital anomalies.

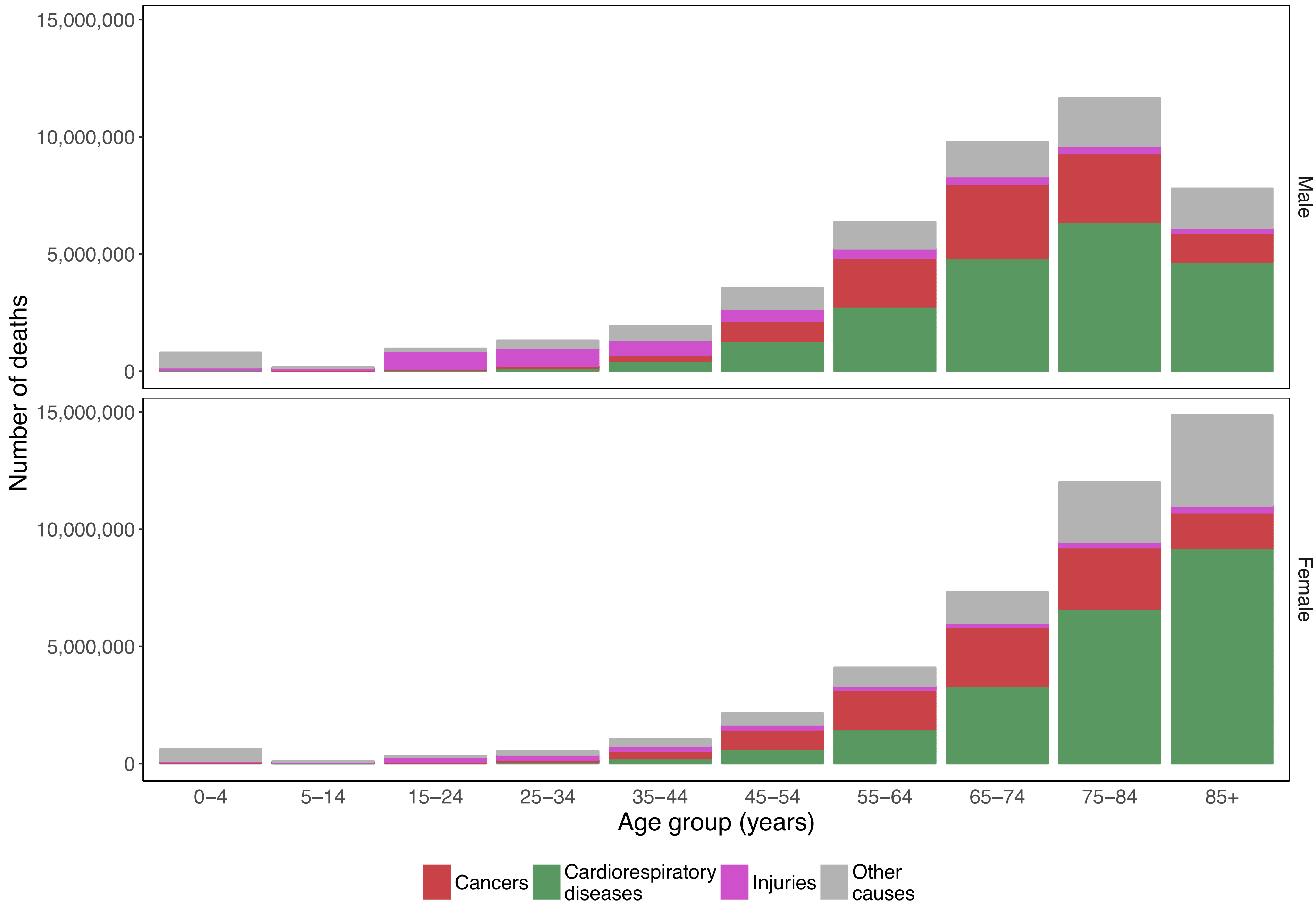

Figure 2. Number of deaths by cause, sex and age group in the contiguous United States for 1980-2016.

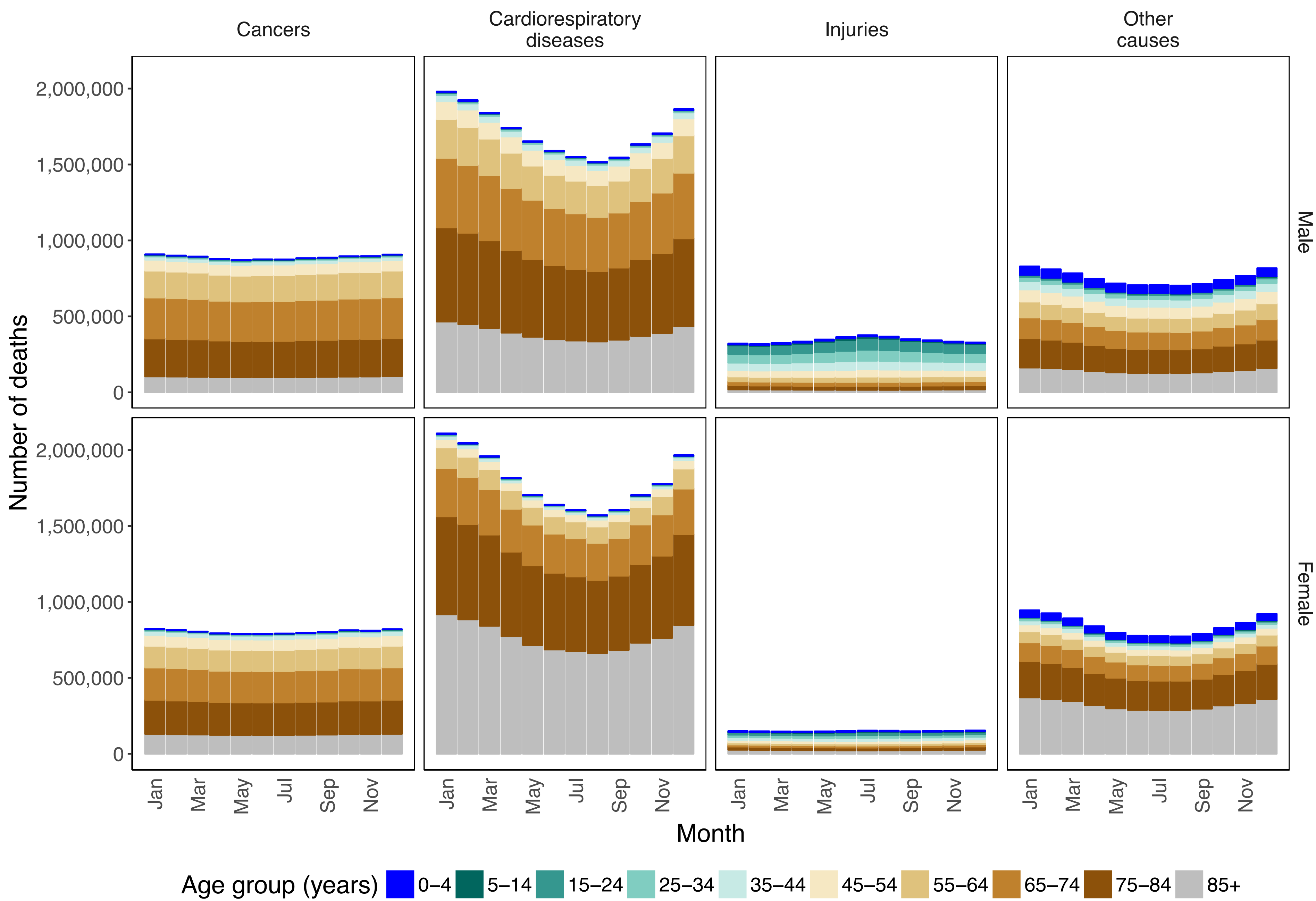

Figure 3. Number of deaths by cause, month, sex and age group in the contiguous United States for 1980-2016.

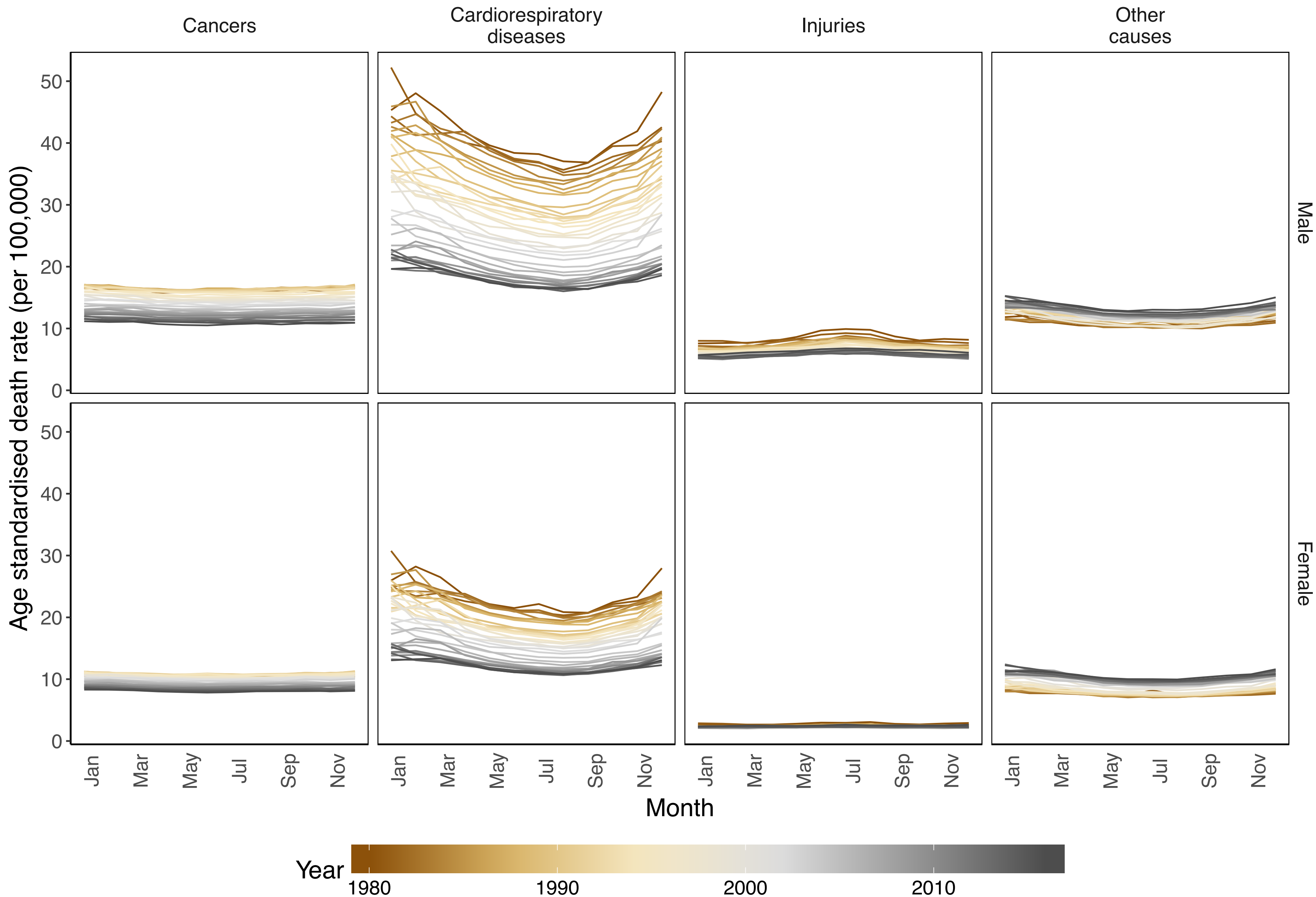

Figure 4. National age-standardised death rates from 1980 to 2016, by cause and month.

Sub-national variation in death rates was also evident. This is illustrated by the state-level summaries of annual age-standardised death rates of cardiorespiratory disease deaths in Figure 5. Figure 5 also shows the variation in change of death rates over time, with death rates in different states changing more rapidly or slowly than other parts and with varying trajectories. This variation by time and space demonstrated that I needed to think about how a statistical model would incorporate sub-national variation, described in Chapter 5. Mortality characteristics of the states for the period of study (1980-2016) are summarised in Table 1.

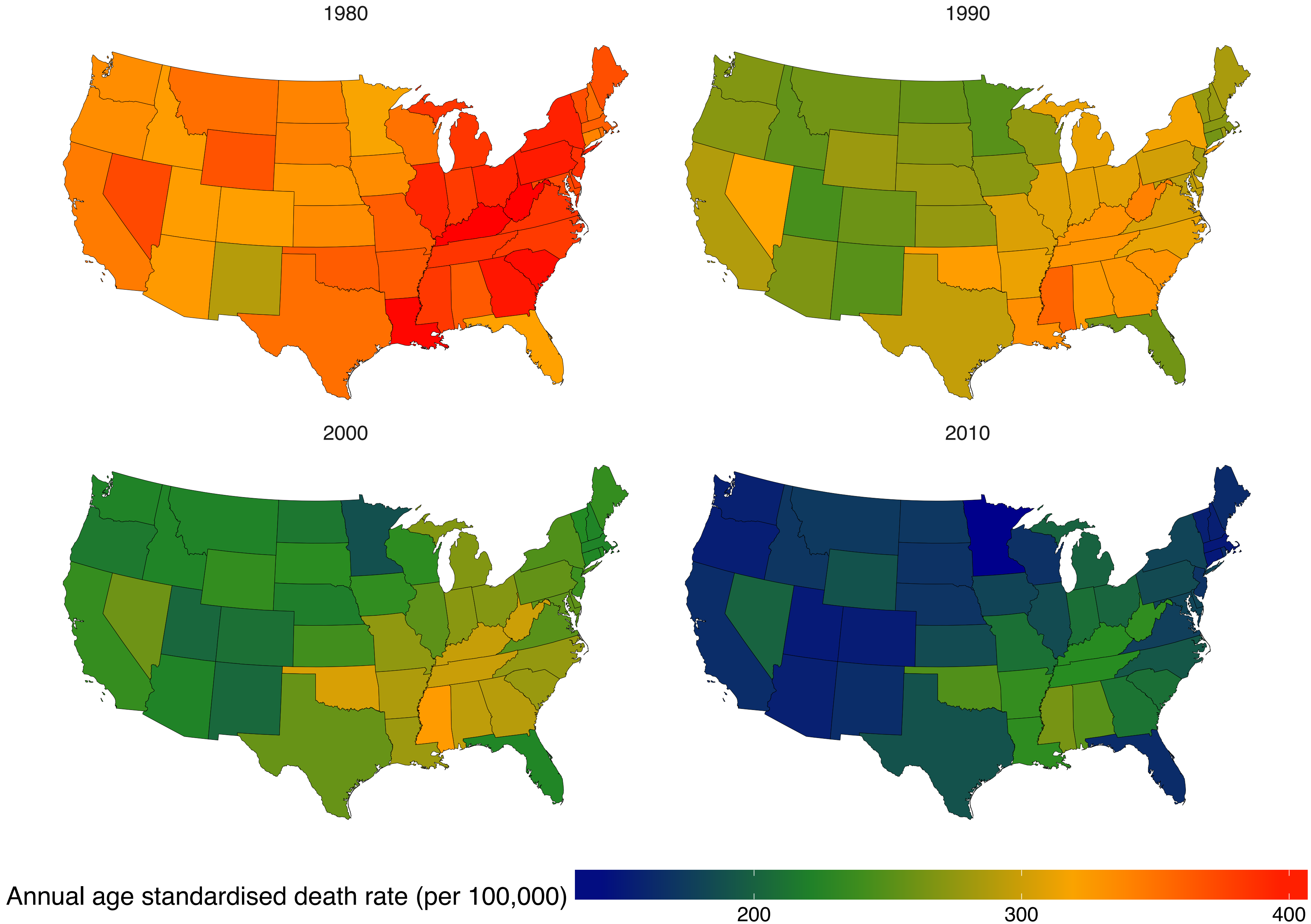

Figure 5. Cardiorespiratory disease annual age-standardised death rates in four example years at the beginning of each decade throughout the study period.

		1980					2016				
	Sex	**Min**	**25th**	**50th**	**75th**	**Max**	**Min**	**25th**	**50th**	**75th**	**Max**
Population	M	242,903	645,187	1,501,479	2,650,750	11,733,600	298,942	950,671	2,289,370	3,641,400	19,493,361
	F	231,478	663,921	1,608,566	2,825,623	1,2059,240	286,559	956,445	2,392,296	3,646,600	19,756,656
Deaths, all causes	M	1,870	5,389	14,481	25,063	99,751	2,514	9,378	21,640	33,661	135,238
	F	1,348	4,479	12,714	21,112	86,763	2,208	8,502	21,301	32,811	1,269,78
Deaths, cancers	M	313	1,118	3,129	5,411	21,301	518	1,852	5,142	8,041	31,547
	F	263	1,037	2,584	4,309	19,354	440	1,721	4,228	6,746	29,394
Cancer deaths as proportion of all deaths (%)	M	16.4	19.8	21.1	21.8	24.2	19.0	22.2	23.0	23.9	25.4
	F	18.1	19.7	21.0	22.2	24.3	17.1	20.5	21.1	22.6	23.6
Deaths, cardiorespiratory diseases	M	959	2,737	8,490	13,931	51,946	990	3,321	8,465	12,746	55,473
	F	754	2,665	7,687	12,141	51,843	862	3,226	8,431	12,802	54,008
Cardiorespiratory disease deaths as proportion of all deaths (%)	M	41.2	52.1	54.5	56.5	59.1	34.9	38.1	39.4	40.7	45.7
	F	47.1	56.6	58.5	59.9	63.2	33.9	38.4	40.1	41.5	45.5
Deaths, injuries	M	240	669	1,698	2,662	13,083	271	1,208	2,220	3,182	12,495
	F	108	288	627	949	4,864	121	623	1,131	1,635	5,241
Injury deaths as proportion of all deaths (%)	M	7.1	9.4	10.4	12.6	20.0	6.8	9.1	9.9	10.8	13.2
	F	3.1	4.1	4.8	5.7	12.1	3.5	4.7	5.2	5.8	7.9
Deaths, other causes	M	254	794	2,180	3,528	13,421	641	2,977	5,922	9,354	35,723
	F	221	708	2,006	3,474	12,311	698	2,887	7,120	11,053	38,335
Other cause deaths as proportion of all deaths (%)	M	11.3	12.8	13.4	14.5	21.4	22.6	25.8	27.6	29.0	33.4
	F	12.3	14.2	15.1	16.7	23.4	26.5	32.1	33.4	34.7	37.8

Table 1. Summary statistics by percentile for states used in the analysis.

3.3 Meteorological data

Here, I describe how I dealt with finding a reliable data source for meteorological records; how I developed an algorithm to process meteorological records for use as an input to my epidemiological model; how I tested results against another dataset; and how my methods are generalisable for use in other studies.

3.3.1 Objectives

My objectives in designing county- and state-level monthly temperature summaries involved taking gridded reanalysis weather data and converting it to United States county-level summaries; and creating population-weighted monthly means and anomalies by United States state.

3.3.2 Sources of meteorological data

Weather stations provide the apparatus to measure in-situ meteorological data. In industrialised nations, the density of weather stations has mostly remained high or increased. For example, in the United Kingdom, the average density of station networks measuring temperature over the past half century is one station per 21 x 21 km^2, with little change throughout the period.[206] In the United States, this value is around one station per 40 x 40 km^2.[207] Particular challenges exist in the developing world, with coverage in Africa remaining generally below a required density to inform research and policy decisions.[208]

Additional measurement platforms such as ships, balloons, buoys, radiosondes, aircraft, scatterometers and satellites have supplemented weather stations to create a rich and live global dataset which are used in weather modelling and prediction.[209] However, while crucial for short-term weather prediction, their use in climate research by themselves is limited due to frequent changes in systems of observation and data assimilation methods.[210,211]

3.3.3 Weather reanalysis data

Creating reliable long-term global climate datasets has become a major national and international concern, partly to help understand the large-scale climate change occurring on Earth. The long-term data are therefore essential inputs into any historical climate record modelling.

Since the 1990s, retrospective analysis, or reanalysis, has provided an increasingly popular technique of assimilating millions of data points of meteorological data from disparate sources into a single model output. This process produces an estimate of the state of the atmosphere at a particular instant in time. Reanalysis of meteorological data consists broadly of three stages: a data quality control stage to ensure high data input standards, a data assimilation module to integrate historical data and an archiver to save the model output.[212]

In the data assimilation stage, a physical model progresses through time in discrete steps while using data from weather measurements to correct and steer the model. The observational and model information is combined using a Bayesian statistical model in which probability distributions are associated with observations and model progress is dictated by physics. There are several basic strategies by which the data are assimilated. These strategies include sequential intermittent assimilation, or three-dimensional variational data assimilation (3DVAR), where a model run is only informed periodically by observations made in the time interval between assimilations. The state-of-the-art data assimilation methodology is four-dimensional variational data assimilation (4DVAR), where a model and the data analysis occur 'live', with observations included at the point in time at which they are made, which improves upon 3DVAR.[213] By assimilating data in this way, output from reanalysis have become ever more trustworthy where there is a good foundation of source data, such as in the United States in the past few decades.

There have been several notable reanalysis developments over time, starting with modelling a handful of atmospheric variables to more than 100. In the first generation, the NCEP/NCAR 40-year reanalysis project, released in the mid-1990s and using 3DVAR assimilation, provided the first global reanalysis at a resolution of around 210km worldwide.[212] The second generation included ERA-40[214] (1957-2002), a reanalysis product developed by the European Centre for Medium-Range Weather Forecasts (ECMWF), with successive improvements in data assimilation using 3DVAR and use of increased available computing power to provide higher resolution. ERA-Interim, released by the ECMWF in 2011 as an 'interim' update to ERA-40,[215] represents the third generation, with an improved atmospheric model and 4DVAR assimilation system. ERA-Interim has global coverage at a 4-times daily frequency across several decades, currently from 1979 to the present day at a resolution of 80km. The latest in line of ECMWF reanalysis is ERA5, released by ECMWF in early 2019, which possesses a 31km spatial and hourly time resolution while also assimilating data using 4DVAR.[216]

3.3.4 Converting gridded meteorological data to area-weighted summaries

I developed an algorithm to convert gridded meteorological reanalysis datasets to population-weighted state summaries using open-source mapping shapefiles along with population data. I used this algorithm to prepare data for my thesis. It has further applicability outside of the bounds of my immediate focus; for example I used the same algorithm to prepare data in research in the impacts of air pollution on life expectancy in the United States.[217]

I summarised the 2-metre temperature estimates available on the ERA-Interim reanalysis. I downloaded each year in my period of study (1980-2016) in the netCDF file form from ERA-Interim website (https://apps.ecmwf.int/datasets/data/interim-full-daily/levtype=sfc/). The file for each year contained data for 2-metre temperature in Kelvin in a global gridded dataset. The grid size, shape and position were consistent throughout the entire period of study. Each grid recorded temperature values at four times per day (00:00, 06:00, 12:00, 18:00) for the entire year. I calculated daily mean temperature for each grid by calculating the mean of the four values in each day. I also obtained the daily maximum and minimum temperatures for each grid by taking the maximum and minimum of the four values in each day. I converted the Kelvin values to Celsius.

A subset of grid squares from the files intersects the contiguous United States. Each of the 3,108 counties in the contiguous United States intersects at least one grid square. I calculated the proportion of each county's area intersecting with the grid squares incident upon it. Using these proportions, I calculated an area-weighted 2-metre temperature value based on grid values for each of the counties for each year's data. This process is illustrated in Figure 6.

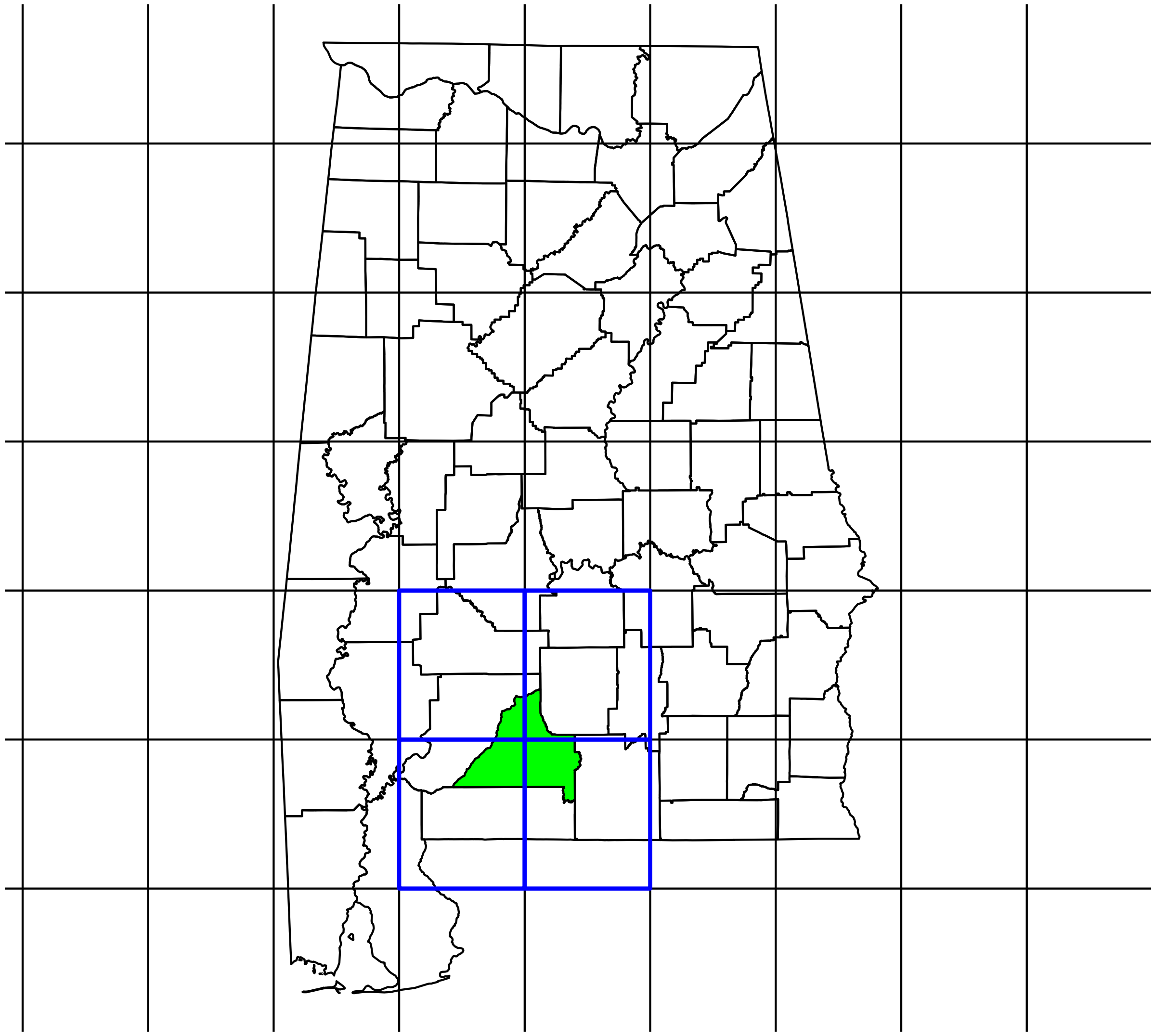

Figure 6. Example of method for generating county-level temperature summaries. Here, the temperature of Conecuh County, Alabama (in green) is generated from a area-weighted average of the four grids which intersect it (in blue).

3.3.5 Creating state-level summaries

The populations of the counties from 1980 to 2016 ranged from 32 (Loving County, Texas) to 10,043,000 (Los Angeles County, California). In small counties, some age groups had zero population in some years, which makes it impossible to calculate death rates. To avoid zero populations, I decided to summarise 2-metre temperature at a state level, which matched the unit of analysis for death rates. Summarising at state level also assures computational tractability for the Bayesian spatio-temporal model, described in Chapter 5.

The population within a state is usually distributed through its counties unevenly, mostly concentrated around urban areas. I needed to take this into consideration when creating state-level temperature summaries. I created a state-wide population-weighted average of 2-metre temperature for each sex and age group for the entire contiguous United States. This would ensure that at the scale of state, the temperature value assigned would reflect the experience of those within the state during that month. I then summarised for a particular month in my period of study in two ways. First, I calculated the county-level monthly averages and made a population-weighted average across the state. Second, I calculated the state population-weighted average by day and averaged across the month. The results were not sensitive to the choice of method. This was also true for generating anomalous temperature values, the concept of which I explain later in this chapter. I used values calculated from county-level averages with a population-weighted average made across the state. I then generated monthly temperature statistics. Figure 7 shows a map of the norm temperatures by state for each month across the contiguous United States during 1980-2016.

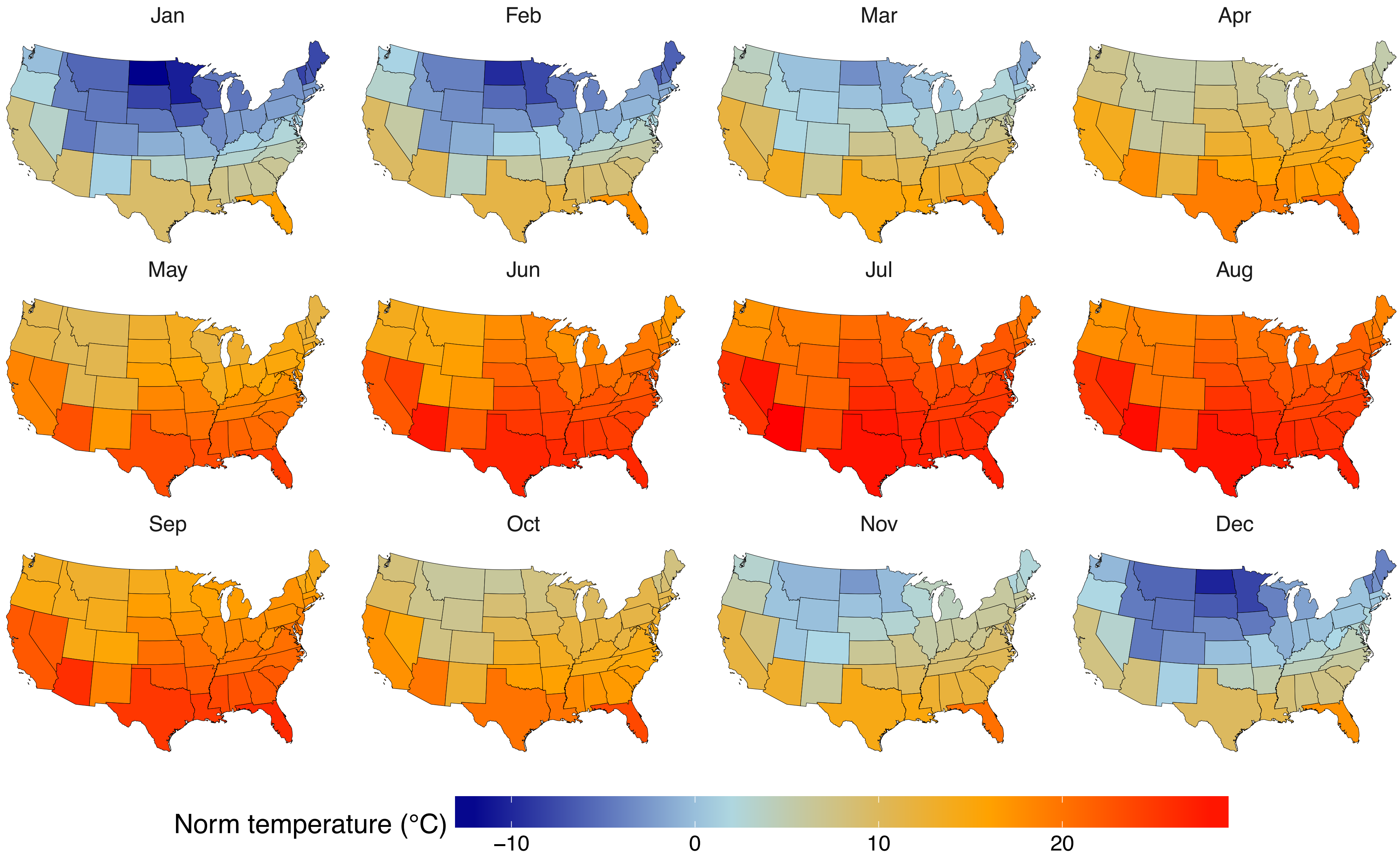

Figure 7. Map of norm temperature (°C) from 1980 to 2016, by state and by month.

As a sensitivity analysis, instead of building monthly temperature values based on daily mean temperatures, I used daily maxima and minima. These measures were strongly correlated to those generated from daily means (Table 2), and therefore I did not run models using these alternatives.

	Mean correlation coefficients between monthly anomaly summary values	
Month	**Mean daily temperature with maximum daily temperature**	**Mean daily temperature with minimum daily temperature**
January	0.98	0.98
February	0.98	0.98
March	0.97	0.97
April	0.97	0.96
May	0.96	0.94
June	0.95	0.92
July	0.97	0.94
August	0.96	0.93
September	0.93	0.91
October	0.91	0.93
November	0.96	0.97
December	0.98	0.92

Table 2. Correlation coefficients between monthly temperatures generated from mean daily and maximum or minimum daily temperatures. Each correlation coefficient was calculated in each state for each month for 1980-2016. The values shown are the means over all states for a particular month.

3.3.6 Anomalous temperature

To mimic the conditions that may arise with global climate change, I developed methodology to examine how deviations from long-term norm temperature may impact death rates. I defined a measure of anomalous temperature for each state and month compared to long-term norm temperature of the state in that month. To calculate the magnitude of temperature anomaly by state and month, I calculated 30-year (long-term) norm temperatures (from 1980-2009) for each month in each state. I calculated for 30 years because it is the duration used in climate assessments.[30] I subtracted these long-term norm temperatures from respective monthly temperature values to generate a temperature anomaly time series for each month and year in each state (Figure 8). The temperature anomaly metric measures the extent that temperature experienced in a specific month, year and state is warmer or cooler than the long-term norm to which the population of each state has acclimatised. These values can be different for neighbouring months in the same state, and neighbouring states in the same month.

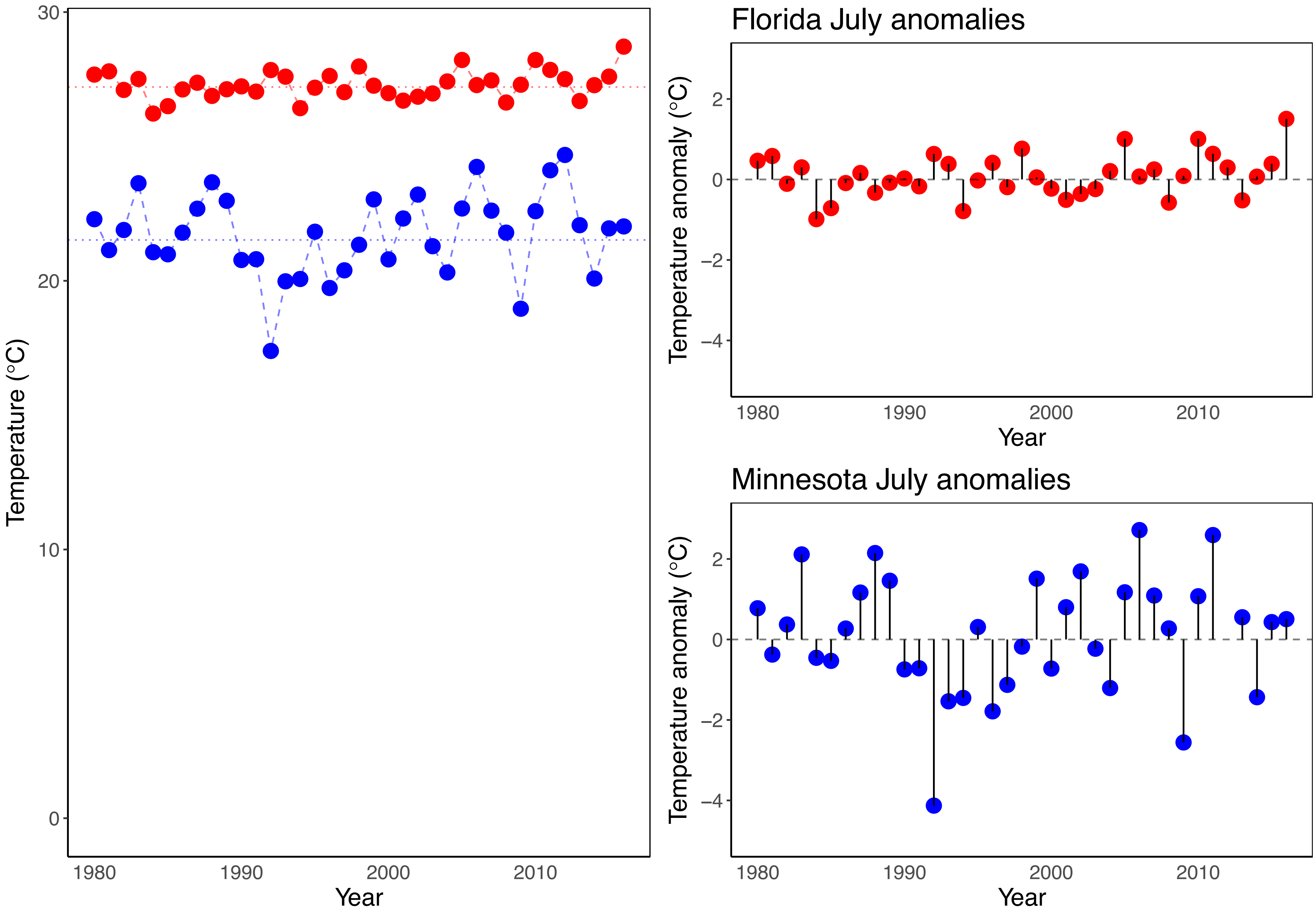

Figure 8. Graphic representation of temperature anomaly measure used in the analysis. The graph shows how monthly temperatures in July two example states (Florida in red and Minnesota in blue) (left panel) for 1980-2016 are used to calculate temperature anomalies. As seen, a warmer state like Florida (top right) can have less inter-annual variation in a particular month (here, July) compared with a cooler state like Minnesota (bottom right).

In this approach, a state with higher, but more stable, temperature in a specific month has smaller anomalies than one with lower but more inter-annually variable temperature. Average size of anomaly over the study period (1980-2016), a measure of how variable temperatures are around their central state-month long-term norm, ranged from 0.4°C for Florida in September, to 3.4°C for North Dakota in February (Figure 9). The size of anomaly had a median value of 1.2°C across all states and months, with 27% less than 1°C and 90% less than 2°C (Figure 9). Temperature anomalies were largest in January and December and smallest in August and September. They were larger in northern and central states than in southern and coastal ones.

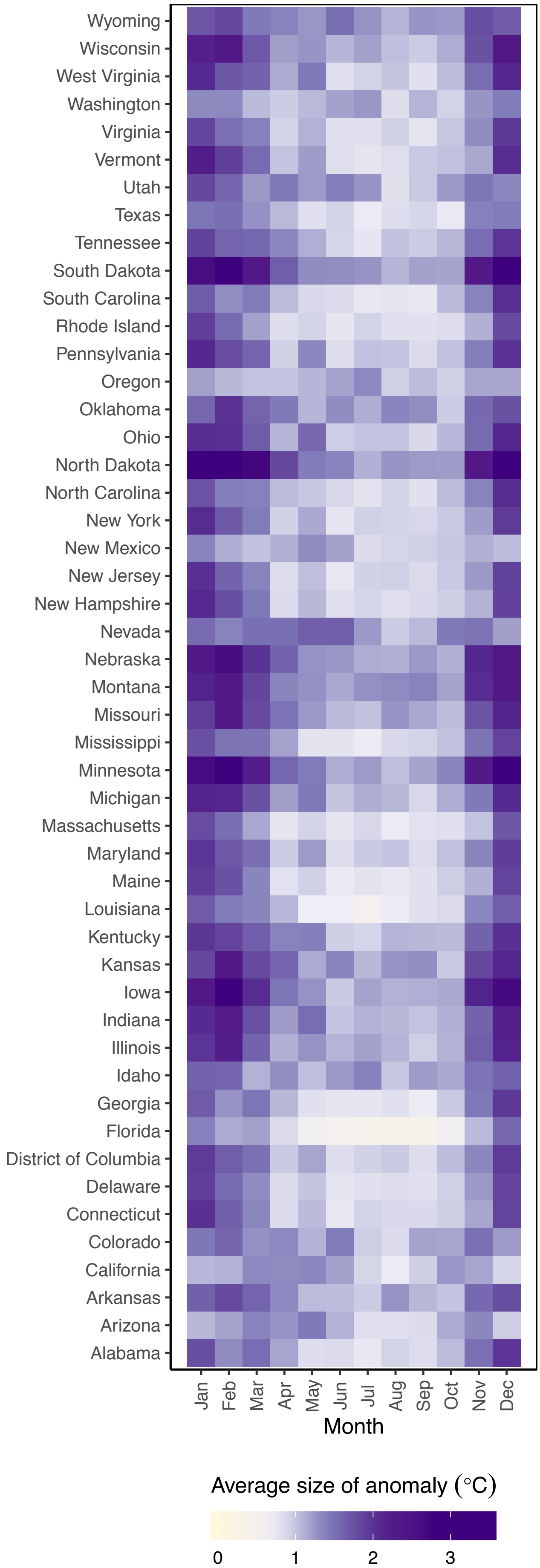

Figure 9. Average size of temperature anomaly (°C) from 1980 to 2016, by state and month. The value for each state and month is the mean of the absolute size of anomaly, be it cold or warm, and hence gives an indication of the scale of anomalies around the norm local temperatures.

3.3.7 Comparing county- with state-level anomalies

I also calculated county-level anomalies to compare how similar they would be to respective state-level ones. I followed the same process for county-level summaries as I did for states in Section 3.3.5. I then compared the county-level anomalies to the respective state-level anomalies over time and calculated the average difference between them (Figure 10). While there were some absolute differences within sets of counties, temperature anomalies in counties generally followed state-wide anomalies, as indicated by the proximity of most points on Figure 10 to the horizontal axes, with 94% an within an absolute deviation of 0.1°C, and 100% within 0.25°C. The largest deviations between county anomaly and state anomaly were typically for counties with lower proportion of population of the entire state. This minimised the impact of misclassification.

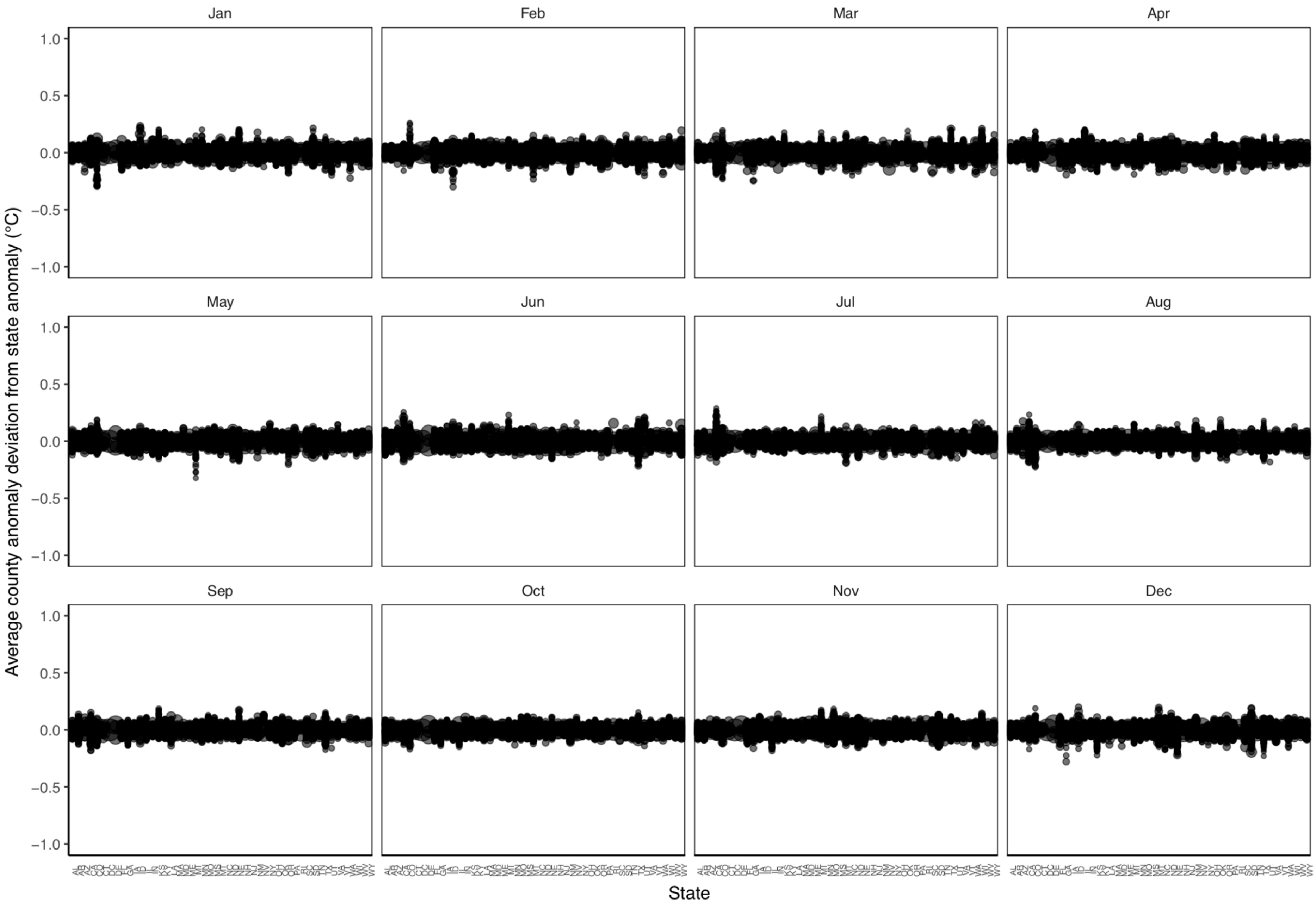

Figure 10. Average county temperature anomaly deviation from state-level anomaly (°C) from 1980 to 2016, by each state and month. Each point represents a county. The size of the point is proportional to its proportion of the respective state's population.

3.3.8 Comparison with PRISM dataset

Any new method of obtaining meteorological summaries should be compared with existing datasets. Other datasets include that from the Parameter-elevation Regressions on Independent Slopes Model (PRISM),[218] which focuses on providing high resolution meteorological datasets for the contiguous United States. In contrast, ERA-Interim provides global coverage, and is therefore a viable dataset if looking for consistent multi-national data sources. Even for the United States, ERA-Interim (and its successor ERA5) would provide additional coverage of non-contiguous states and other American off-shore insular areas, such as Hawaii, Alaska, Puerto Rico and Guam.

I compared the results of processing gridded ERA-Interim reanalysis data to those of the PRISM data. An example of the correlation between monthly temperature summaries for each grid square in the contiguous United States is demonstrated in Figure 11 (correlation coefficient 0.99). Other years and months in isolation were similarly correlated. The method of calculating temperature in PRISM was also different for the first year of my study, 1980, compared with 1981 onwards. This contrasted with ERA-Interim, which was consistent with the methodology and data assimilation across the entire period.