

Time series design and statistical modelling health essay

[Health & Medicine](#)



Significant associations between heat stress and health outcomes have been reported in several epidemiological studies . Those studies have identified these associations by using different heat stress variables or statistical modellings. A number of studies consider deaths above the threshold temperature , deaths at particular percentile of temperature or taking into account of the short-term effects and harvesting effects . Also, some studies explored the relationship between mortality and temperatures alone or effect modification by certain factors such as high humidity, wind speed, air pressure and cloud cover, as well as the possible synergistic effects from air pollution . Statistical modelling for heat-related deaths Understanding the strengths and weaknesses of various types of studies will help determine the suitable approach for identifying the association between heat stress and mortality in this thesis. This section summarises the statistical approaches often used in the studies on heat-related deaths including time-series and case-crossover studies.

Time-series design and statistical modelling

In environmental epidemiology studies, the time-series approach is a well-known study design that documents the association between temperature and mortality by using a wide range of modelling techniques. The typically time-series regression is suitable for estimating the effects of heat stress exposure over the period of time from one day to several days by comparing daily counts of deaths as an outcome variable with the temperature on the same day or over the last few days (or lagged days) . This time-series design of heat-related deaths aims to estimate the change in the number of deaths associated with temperature, controlling for potential confounders such as

long-term trend, season, air pollution, or epidemics of influenza . Similar to the study methods for air pollution epidemiology, the time-series approach is often used for quantifying an association between air pollution and daily death. However, there are some differences between the time-series study for assessing the health impacts from air temperature and the heat impacts from air pollution. For example, the temperature-related mortality has a non-linear relationship such as a U- or V-shape , while the air pollution-related mortality has a linear association . The statistical modelling techniques for time-series studies have been developed from simple linear regression methods to more sophisticated modelling approaches. The daily counts of deaths can be assumed to be a Poisson distribution thus the simple methods of Poisson regression models can filter out the long-term trend (such as population growth or change and technology change) and seasonal variation in daily health outcomes and weather variation . The techniques for conducting non-linear regression analysis in time-series studies such as generalised linear model (GLM), generalised additive model (GAM) or generalised estimating equation (GEE) are commonly used in the recent time-series analyses . These three models allow users to adjust for linear or non-linear terms of daily temperature, over-dispersion or confounding effects of trends and seasonality. Some studies used the GLM with parametric regression splines such as natural and cubic splines. A single spline function of time has been used to produce an irregular seasonal trend which is believed to control additional confounding effects operating at medium timescales. The GAM extensively is used for exploring the short-term effects of heat stress on mortality . The GAM is more flexible than GLM because GAM

can be used to explore many nonparametric relationships simultaneously by using nonparametric smoothing such as Loess smoothing technique adjusting for temperature or potential confounders such as season, effects of the day of week, other weather variables and air pollution variables that vary on a daily basis. The dependent variable in GAM is the natural logarithm of the expected number of deaths, and the regression coefficients are the natural logarithms of the rate ratio. However, GAM has a collinearity problem for non-linear functions as well as the difficulty in specifying the type of smoother and the number of degrees of freedom to adjust for changes overtime . Also, the nonparametric smoothing in GAM could lead to biased estimates and to underestimate the true variance . The GEE is an extended approach from GLM . The GEE has been used in the analysis of health effects from air pollution by Schwartz and Dockery . Some recent studies used GEE to investigate the health effects from heat waves . For example, a study by Baccini et al. used the GEE approach to explore the exposure-response curves of heat-related deaths in 15 European cities. They stated that GEE is suitable for longitudinal data analysis when mortality is assumed to be independent of different summer periods, while the deaths in the same summer season are assumed to be correlated .

Case-crossover design and statistical modelling

A case-crossover design has been recently conducted to evaluate the acute effects of temperature on health outcomes . It is adapted from a case-control design to study the effects of transient exposure on the occurrence of acute events . For example, only deaths (cases) during heat wave are analysed by comparing the heat exposure on the day of death with the exposures of

<https://assignbuster.com/time-series-design-and-statistical-modelling-health-essay/>

same person in nearby days. This matched sets of exposure in the same person are called "reference windows". The case-crossover design is suited for exploring the temperature-related deaths when individual-level mortality records data are available. Date record of person death is considered as a case day comparing with nearby days (control days) in order to represent the exposure distribution during the time periods when that person is alive. There are many different types of case-crossover designs depending on the selection of control days. Unidirectional design, bidirectional designs (and their subtypes) and up to time-stratified case-crossover designs have been developed respectively. First, the unidirectional design is the initial case-crossover design. This design selects one or more control days before the case day. However, the unidirectional design does not control for the trends over time of exposure variables or health outcomes, and so can be subject to bias. Therefore, if there is no trend in exposures, a unidirectional design can be used by selecting control days prior to the case day. The bidirectional designs have been developed to reduce the bias of time trends in exposure which there are three different subtypes in the choice of control periods. The different types in bidirectional case-crossover designs include the full-stratum case-crossover, symmetric case-crossover and semi-symmetric case-crossover design. First, the full-stratum bidirectional case-crossover is designed to choose the control days from all exposure periods in the time series before and after the case days. This full-stratum bidirectional design controls time trends in exposure, but does not control seasonal patterns in exposure or health outcomes. Second, the symmetrical bidirectional case-crossover design takes two control days including one before and one after

the case day and those control days far equally from the case day . This method can control seasonal trends which occur in the above design. However, there is the potential for selection bias because the cases at the beginning or the end of the data series have fewer control days for matching . Navidi and Weinhandl noted that the symmetric case–crossover design is biased by time trends of exposure. For example, a study by Liu et al. used the symmetrical bidirectional case-crossover design to study effects of heat waves on daily death counts in Beijing, China by selected the seventh day before and after death as its own self-control between 1 January 1999 and 30 June 2000. They found a significant increase in respiratory mortality during heat wave with the highest impacts among females aged older than 65 years . Third, the semi-symmetric case–crossover design selects a control day only before or after the case day at random . This design can control the long-term and seasonal trends the same as the symmetric bidirectional design. Nevertheless, the semisymmetric bidirectional design selects only one control day at a fixed interval, so the estimates can be biased . However, the previous case-crossover designs have problem as a result of inappropriate selection for the control days in non-disjointed strata, this bias called the " overlapping bias" . In order to reduce the bias that could appear in above designs, the time-stratified case–crossover design was developed . Janes et al. demonstrated that the time-stratified design can avoid the overlapping bias by selecting one or several control days falling within the same time stratum when the case events occur. For example, the control days can be matching on the same day of the week as the case within the same month to create partitions with 3 or 4 control days for each case. This

stratified by month and by day of the week gives the same maximum distance between case and control. This method can avoid autocorrelation problem and control potential confounders such as day of the week by adding to the model as indicator variables . Advantages of the case-crossover design, it can eliminate the effects of potential confounders by matching methods, and can examine whether the events are associated with a particular exposure . Using matched data and comparison within the same person can control confounding effectively. This study design automatically adjusts for personal confounding such as age, socioeconomic status, personal smoking or dietary habits, as they remain constantly for each person during case and control periods . Moreover, it can reduce bias from selection and misclassification of exposure, overmatching. Also, it can adjust for time-trends and seasonal variation . Regarding the effects of a long- and short-term trend of seasonality in daily data, the control days should be selected close to the case day . The measure of association from case-crossover study can be shown by odds ratio (OR) or relative risk (RR) and 95% confident interval (95% CI) . Typically, the OR is estimated by using conditional logistic regressions or stratified Cox proportional hazards models . The conditional logistic regressions are often used to estimate adjusted ORs by dividing the number of cases exposed during the case day by those exposed during the control days. Basu et al. identified the association between apparent temperature and preterm delivery by using a time-stratified case-crossover analysis in California during the summer months from 1999 to 2006. The control days were selected at lag six of apparent temperature exposure. They found the association between

preterm deaths and the weekly average apparent temperature . Also, Basu et al. used the time-stratified case-crossover design to explore the association between temperature and cardio-respiratory mortality among the elderly population in the United States. Furthermore, they compared results from the case-crossover design to the time-series analyses. The results showed that the case-crossover design and the time-series analyses are comparable as they provided consistent findings in this study . The similar results from the case-crossover design and the time-series log-linear model without over-dispersion was also reported in a study by Lepeule et al. . However, bias in the estimation of health effects by using the time-stratified case-crossover design can occur when the health outcomes and exposure have a trend . For example, the time-stratified case-crossover design cannot capture the smoothness of daily death under the seasonal trends or day of week effects . Lu et al. suggested that the future study using the time-stratified case-crossover design should do the pre-model checking for the time trends to make sure whether they are constant during case and control periods

Threshold temperature and shape of association

The outcomes from the above quantitative studies of heat mortality relationship commonly reported a regression coefficient, relative risk from time-series analysis, odds ratio from case-crossover design, or percent change in deaths . These results can show the magnitude of the heat stress effects when the temperature increase at certain level. In addition, studies on the relationship between heat stress and mortality should involve the complexity of heat exposure and body mechanism through an investigation the dose-response shape of heat and mortality and an exploration of the temperature threshold. The threshold

temperature is defined as a temperature with no or minimum effect on mortality. It is commonly used to estimate the health effects from exposure to high or low temperature . If the threshold temperature exists along with the exposure-response relationship, it will present the non-linear association between temperature and health outcomes as in U-, V- or J-shaped relationship . In some countries, there is a U- or V-shaped relationship between temperature and mortality with the lowest number of deaths at bottom of the U-shape. There is an increase in death during hot weather as well as cold weather . Some studies described the relationship between daily mortality and daily temperature as in reverse J-shaped pattern . Where the trough of the J-shape represents the comfort zone and the short arm of the J-shape represents the steeper slope of mortality increase with rising temperature above this comfort zone. The long arm of the J-shape shows the increase in mortality at colder temperatures below that comfort zone . Few studies reported that cities with warmer climates have higher threshold temperatures than cooler cities . These findings confirmed that the threshold temperatures vary by geography location, weather condition and population group. Consequently, the estimation of heat-related death should take into account regional, seasonal and demographical differences . Lagged effect of heat exposureThe risk of mortality from heat stress exposure can occur on the current day and also persist to several days later . The distribution of cumulative effects of heat stress exposure over days or weeks after exposure has been addressed in many time-series studies. Some studies applied the distributed lag models (DLMs) to create the lag structure of exposure-response relationships . This distributed lag model is suitable to

estimate the heat stress effects particularly during heat waves periods as it allows the linear effect of single heat exposure to be distributed over those period of time . Recently, the distributed lag models have been developed from simple distributed lag models to distributed lag non-linear models (DLNMs) . The DLNMs are flexible to study the delayed effects and non-linear association between temperature and mortality . Many studies about temperature-related death showed that the longer lag times are common at cold weather, and shorter lag times are often at hot weather . As a result, the deaths from high temperature are likely to happen quickly while deaths from low temperature can occur several days after the cold weather. The longer lag times also showed a deficit in mortality if there is mortality displacement associated with the heat stress effects . Mortality displacement or short-term harvesting of death is the key question for assessing impacts of heat stress on mortality in the time-series studies and particularly the studies of the number of deaths in extreme events such as heat waves . This is very important for policy makers to use the evidence-based information of heat stress effects on deaths by taking into account the mortality displacement issues . During heat waves, the mortality displacement refers to the extreme temperature events that induce a short-term forward shift in the observed mortality distribution among selected population, or in some locations. The deaths increase during the extreme temperature and then decrease in a few days or weeks later regardless of high heat exposure . Here, the mortality displacement affects people who are very near death and have been brought forward by the heat stress effects within a short period of time such as elderly people or those who

have chronic diseases . Therefore, the investigation of the heat wave effects on mortality in the time-series analyses should involve the harvesting effects that reduce the overall long-term impacts of heat stress, and should take in to account of the mortality displacement issue in order to produce the accurate estimates .