OPEN

# Efficiency of case-crossover versus time-series study designs for extreme heat exposures

Caleb Schimke<sup>®a,\*</sup>, Erika Garcia<sup>®a</sup>, Sam J. Silva<sup>a,b</sup>, Sandrah P. Eckel<sup>a</sup>

**Background:** Time-stratified case-crossover (CC) and Poisson time series (TS) are two popular methods for relating acute health outcomes to time-varying ubiquitous environmental exposures. Our aim is to compare the performance of these methods in estimating associations with rare, extreme heat exposures and mortality—an increasingly relevant exposure in our changing climate.

**Methods:** Daily mortality data were simulated in various scenarios similar to observed Los Angeles County data from 2014 to 2019 (N = 367,712 deaths). We treated observed temperature as either a continuous or dichotomized variable and controlled for day of week and a smooth function of time. Five temperature dichotomization cutoffs between the 80th and 99th percentile were chosen to investigate the effects of extreme heat events. In each of 10,000 simulations, the CC and several TS models with varying degrees of freedom for time were fit to the data. We reported bias, variance, and relative efficiency (ratio of variance for a "reference" TS method to variance of another method) of temperature association estimates.

**Results:** CC estimates had larger uncertainty than TS methods, with the relative efficiency of CC ranging from 91% under the 80th percentile cutoff to 80% under the 99th percentile cutoff. As previously reported, methods best capturing data-generating time trends generally had the least bias. Additionally, TS estimates for observed Los Angeles data were larger with less uncertainty.

**Conclusions:** We provided new evidence that, compared with TS, CC has increasingly poor efficiency for rarer exposures in ecological study settings with shared, regional exposures, regardless of underlying time trends. Analysts should consider these results when applying either TS or CC methods.

Keywords: Climate change; Statistical methods; Heat-related mortality; Climate and health; Simulation

### Introduction

Under climate change, it is crucial to understand the health effects of extreme weather events, which are occurring more frequently and increasing in intensity. Heat records were broken in all continents in 2022, the year with the highest global temperatures in 100,000 years.<sup>1</sup> Extreme heat events produce extensive and well-documented excess mortality, as in the case of the 2003

<sup>a</sup>Department of Population and Public Health Sciences, Keck School of Medicine, University of Southern California, Los Angeles, California; and <sup>b</sup>Department of Earth Sciences, Dornsife College of Letters, Arts and Sciences, University of Southern California, Los Angeles, California

Supported by the Southern California Environmental Health Sciences Center, National Institute of Environmental Health Sciences grants P30ES007048 and T32ES013678.

The observed LA County mortality data are not to be made available because of California state policies. All code and one simulation of mortality will be made available in a repository on Zenodo, accessible via a unique Digital Object Identifier (https://doi.org/10.5281/zenodo.14630196).

**SDC** Supplemental digital content is available through direct URL citations in the HTML and PDF versions of this article (www.environepidem.com).

\*Corresponding Author. Address: Department of Population and Public Health Sciences, Keck School of Medicine, University of Southern California, 1845N. Soto Street, Los Angeles, CA 90032. E-mail: schimke@usc.edu (C. Schimke).

Copyright © 2025 The Authors. Published by Wolters Kluwer Health, Inc. on behalf of The Environmental Epidemiology. All rights reserved. This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

Environmental Epidemiology (2025) 9:e370

Received 10 October, 2024; Accepted 15 January, 2025

Published online 13 February 2025

DOI: 10.1097/EE9.000000000000370

heatwave in Europe and North Africa.<sup>2,3</sup> Much of the statistical methodology for studying the association of exposures such as air temperature with various health effects have been adopted from the well-established field of air pollution epidemiology. However, the performance of these methods for extreme and rare exposures has not been fully explored.

Exposure to ambient air pollution, with informative day-today variability, is different from exposure to extreme weather, which happens infrequently by definition. Here, we describe two methods that have been adapted from the field of air pollution epidemiology and used to study the health effects of extreme weather exposures. Ecological time-series (TS) analysis with Poisson regression (denoted henceforth as TS) relates transient exposures to total counts of acute outcomes collected at regular intervals for a defined population. Because the location(s) serves as its own control, there is no need to control for timeconstant confounders. Control for time-varying confounders can be achieved by including a smooth function of time (e.g., using a regression spline like a natural cubic spline) in the Poisson regression model.4,5 A crucial and greatly influential decision in these models is how flexible to make these smooths (i.e., choosing the number of degrees of freedom [df]), since undersmoothing can result in residual confounding for time-varying confounders and oversmoothing can blunt associations with the

## What this study adds:

This manuscript describes a simulation study based on a real data example that compares two popular statistical methods for estimating the population mortality effect in the context of extreme or rare exposures. This type of analysis is becoming increasingly relevant as climate change literature seeks to quantify the effect of rare, extreme weather events using existing epidemiologic methodologies. Choosing the most appropriate method for the task at hand is an important part of research on how to improve public health and inform policy related to extreme weather events and climate change. time-varying exposure.<sup>6</sup> Case-crossover (CC) studies relate transient exposures within a short-term referent window to an acute outcome on a participant-by-participant basis. Time-invariant factors, such as age, gender, and health status, are controlled for by only making within-participant comparisons.7 Time-stratified CC studies with conditional logistic regression controls for time trends by design through selecting each participant's reference window according to which day of week (dow), month, and year their outcome occurred (i.e., all Mondays during January 2015). Other CC study designs define the reference window differently, such as the symmetric bidirectional (SB) design that selects two or more equidistant time points on either side of the index day. While CC designs are naturally extendable to scenarios in which participant-level data are available, TS models using individualized data, such as case TS, are not as easily implemented and have not been explored in this context. TS and CC methods have been compared in other contexts, particularly in air pollution research,<sup>8,9</sup> but not in the context of extreme weather events. There are concerns about CC designs being less efficient, resulting in more uncertain estimates as compared with TS, with a prior report showing that the SB CC design had 66% of the efficiency of TS under certain scenarios.9 There is a further concern that the problem of reduced efficiency could be amplified for binary indicators of extreme weather in CC studies. where case and control periods with identical exposure values drop out of the traditional estimation approach of conditional logistic regression, providing no information to the analysis.<sup>10</sup>

In this article, we evaluate and compare the performance of TS and CC methods in estimating the association of extreme heat with health in a simulation study and using observed data. We selected extreme heat as an example of "extreme weather" because it is one of the most well-studied extreme exposures in health studies.<sup>11</sup> Using observed daily temperature data and associations with mortality in Los Angeles (LA) County in the years 2015–2016 as our motivation, we generated synthetic mortality data from observed temperature data under several data-generating scenarios. In particular, we dichotomized temperature at the 80th to 99th percentiles to evaluate the impact of increasingly rare exposure extremes. As these rare events become more frequent and intense in our changing climate, efficiently estimating their public health impact using data available today will become increasingly important.

#### Methods

#### Mortality data

Death certificate data for all causes of death between 1 January 2014 and 31 December 2019 were obtained from the California Department of Public Health's Vital Statistics. A TS of daily total all-cause mortality in LA County was produced by summing the number of deaths among decedents with LA listed as their county of residence on each day. The use of these data was approved by the Committee for the Protection of Human Subjects of California (California Health and Human Services Agency's Federalwide Assurance #00000681).

#### Temperature data

Daily 1-hour minimum temperature (Tmin, °C) data were obtained from the GridMet spatiotemporal reanalysis model of meteorological data, which estimates Tmin at a spatial resolution of  $4 \times 4$  km grid.<sup>12,13</sup> The daily Tmin exposure data used in this study were calculated by taking the aggregate mean Tmin across all census tracts within LA County. Based on the distribution of daily Tmin in LA County during the study period, we defined extreme heat days as those in the top 80th, 90th, 95th, 97.5th, or 99th percentiles, similar to prior literature.<sup>12,14</sup> In total, we considered six different versions of the temperature

variable: continuous Tmin and the five increasingly rare binary variables, referred to generically as  $x_t$ .

#### Statistical methods

1

For TS analyses, Poisson regression was used to relate mortality counts on the day t,  $y_t$  (mean  $\mu_t$ ), to that day's temperature,  $x_t$ , while controlling for time-varying confounders as follows:

 $\log(\mu_t) = \alpha + \beta x_t + s(t, df) + dow, where y_t \sim Poisson(\mu_t)$ . (1)

We controlled for seasonality/long-term time trends using a natural cubic spline, s(), of time with a defined number of df as well as for indicators for dow. Estimates of  $\beta$  quantify the log-relative risk of mortality associated with a 1 °C higher daily Tmin or for dichotomized temperature days with "high" versus "low" Tmin.

For CC analyses, the time-stratified referent scheme was used as our primary approach, which has become widely adopted over other schemes, such as the SB scheme.<sup>7</sup> In the time-stratified CC approach, for the participant *i*, the date of death is the case day, and all dates on the same dow and within the same month and year as the case day are selected as control days. The referent window contains one case day ( $y_{ij} = 1$ ) and three or four control days ( $y_{ij} = 0$ ), depending on the month. Traditionally, conditional logistic regression is then used to model the association between the temperatures and binary outcome status on case and control days within the referent window, conditional on referent windows  $\xi_i$  as follows:

$$ogit(P(y_{ij} = 1 | x_{ij}, ..., \xi_i)) = \xi_i + \beta_1 x_{i,j}.$$
(2)

No adjustments for dow or long-term trends are necessary because they are controlled for by design.<sup>7</sup> In practice, data management and computation for this approach are computationally intensive and necessitate an expanded format dataset with *y*<sub>t</sub> number of rows for each case and referent day. We will denote this method as CC<sub>ex</sub>. Equivalent alternative approaches have been identified, including conditional logistic regression using weights to represent multiple cases on the same day (denoted CC<sub>wtd</sub>)<sup>15,16</sup> as well as a conditional Poisson regression model in the TS format (denoted CC<sub>CP</sub>):

$$\log(\mu_t) = \alpha + \beta x_t + \text{stratum}, \text{ where } y_t \sim \text{Poisson}(\mu_t),$$
 (3)

where stratum indicates a unique intercept for each referent window that are conditioned out as nuisance parameters using the "eliminate" argument of the gnm function in the gnm package of the R programming language (R Foundation for Statistical Computing, Vienna, Austria).<sup>16</sup> Advantages of this model include not just the improved computational efficiency and speed but also the ability to calculate the overdispersion parameter using quasi-Poisson regression and the ability to access model-checking tools not available to most conditional logistic regression functions. We used  $CC_{cP}$  as the primary CC method in this work and demonstrated that results from  $CC_{ex}$  and  $CC_{wtd}$  are equivalent (Table S3; http://links.lww.com/EE/A328). Finally, to facilitate a direct comparison of our work with the prior literature,9 we used an earlier version of the CC referent selection strategy, the SB design, which selects control days symmetrically in multiples of 7 days on either side of the case day. We evaluated both a 1-week (CC<sub>SB1</sub>) as well as a 1- and 2-week SB (CC<sub>SB2</sub>) (Table S3; http://links.lww.com/EE/A328). Example R code for performing CC<sub>CP</sub> and TS models is provided in supplementary material (R Codes; http://links.lww.com/ EE/A328).

# Simulation study based on a real data example

#### Data-generating scenarios

Expected daily mortality was generated based on observed daily temperature and mortality associations in LA County (Figure 1)



Figure 1. Daily minimum temperature (A) and daily mortality counts (b) in LA County, California from 1 January 2014 to 31 December 2019. Triangles mark days at or above the 99th percentile of daily minimum temperatures. Predicted daily mortality from two models are plotted with different colors: TS with a natural cubic spline of 8 degrees of freedom (black) and time-stratified CC (gray). Plotted values are for a reference day of the week (Sunday) for clarity of visualization.

under two time trend scenarios: smooth (a la TS, black line in Figure 1) and stratification by month and dow (a la timestratified CC, gray line in Figure 1). For the smooth time trend, a TS model (Eq'n 1) was fit to the observed LA County mortality and temperatures, controlling for time with a natural cubic spline for time having between 1 and 12 df/year, based on which value minimized the Bayesian Information Criterion (BIC). For the stratification by month and dow time trend, a time-stratified CC model was fit to the observed LA County mortality and temperatures (Eq'n 3). Model fitting for each scenario was repeated six times each, one for each version of  $x_t$ . For the TS models on observed LA County data, 8 df minimized BIC for each version of  $x_t$ . Additional adjustment for an indicator of holidays had little impact on the estimated association from the model for observed data (results not shown). The estimated association between observed temperature and mortality,  $\beta$ , for the TS models ranged from 0.004 for continuous Tmin to 0.066 for the 99th percentile dichotomization of Tmin (Table 1). In our primary simulation study, we generated expected mortality counts,  $E(y_t)$ , using Eq'n 1 or Eq'n 3 by holding  $\beta$  fixed at 0.04 for all six versions of  $x_t$  but using values for the other regression coefficients (i.e., spline, dow, and stratum terms) based on what was estimated in the observed LA County data for each scenario and version of  $x_t$ . Hence, we held the magnitude of the true association constant but varied how rare the temperature exposure was. In secondary analyses, expected mortality counts were generated using the unaltered values of  $\beta$  listed in Table 1 that correspond to each version of  $x_t$ , allowing the exposure to become both more rare and to have larger associations. In summary, expected mortality was calculated under four data-generating scenarios for each version of  $x_t$ : TS model with  $\beta$  fixed at 0.04 for all  $x_t$ 

Table 1.		
Application	of methods to observed data from LA county	

	Time series, 8 df/year (TS <sub>8</sub> )		Time-stratified case-crossover (CC <sub>cP</sub> )		
Temperature variable	$\hat{eta}$	SE	$\hat{eta}$	SE	Relative efficiency
Continuous	0.0044	0.0006	0.0026	0.0007	0.90
80th	0.0259	0.0068	0.0173	0.0072	0.89
90th	0.0298	0.0071	0.0190	0.0078	0.83
95th	0.0487	0.0087	0.0384	0.0094	0.86
97.5th	0.0463	0.0115	0.0357	0.0127	0.82
99th	0.0662	0.0172	0.0560	0.0193	0.80

Estimated associations of daily minimum temperature (Tmin), defined as continuous Tmin or a percentile-based dichotomization (*x*<sub>t</sub>), with daily all-cause mortality counts in LA County, CA from 1 January 2014 to 31 December 2019. Relative efficiency is calculated according to  $SE_{TSR}^2/SE_{CC}^2$ .

(Scenario 1); time-stratified CC model with  $\beta$  fixed at 0.04 for all  $x_t$  (Scenario 2); TS model with a different  $\beta$  for each  $x_t$  (Scenario 3); and time-stratified CC model with a different  $\beta$  for each  $x_t$  (Scenario 4). Our primary simulation study used Scenarios 1 and 2 and our secondary simulation study used Scenarios 3 and 4. Expected daily mortality counts were input as the mean  $E(y_t)$  of a random Poisson distribution generator to generate synthetic daily mortality for each scenario and version of  $x_t$ . This method for generating simulated data enables us to preserve realistic, complex exposure patterns while manipulating the data so that the underlying effects are known.<sup>17-19</sup> We also fit models to the original data as an example implementation in a real-life study.

#### Methods considered

We evaluated the performance of the following methods: TS models with a natural cubic spline having 8 df/year (TS<sub>8</sub>, equal to df used to generate simulated data in Scenarios 1 and 3), 4 df/year (TS<sub>4</sub>, an underfitted model), or a dynamically chosen number of df/year that minimizes BIC (TS<sub>Dyn</sub>, mimics the real-life situation where df is estimated from the data)<sup>4</sup>; time-stratified CC models using a computationally efficient method (CC<sub>CP</sub>) as well as supplementary analysis methods using more traditional implementations (CC<sub>wtd</sub> and CC<sub>ex</sub>, described in the Statistical Methods section).<sup>20</sup> Also included in supplementary simulation study results are those from a SB referent scheme,<sup>9</sup> including 1-week (CC<sub>SB1</sub>) or 1- and 2-week (CC<sub>SB2</sub>) models. In summary, the primary simulation study used four methods: TS<sub>8</sub>, TS<sub>4</sub>, TS<sub>Dyn</sub>, and CC<sub>CP</sub>.

#### Number of simulated datasets

For each data-generating scenario and version of  $x_t$ , 10,000 datasets were generated. This number is sufficient to estimate all relevant performance measures (bias, coverage, power, model variance) with Monte Carlo standard error (SE) reduced to a negligible magnitude. Random seeds were saved for internal reproducibility. The first 10 rows of an example simulated dataset can be found in Table S1; http://links.lww.com/EE/ A328. Secondary/sensitivity analyses using the models CC<sub>wtd</sub>, CC<sub>SB1</sub>, and CC<sub>SB2</sub> were tested in 100 simulations for Scenario 1 only (Table S3; http://links.lww.com/EE/A328).

#### Evaluation

Each method was used to estimate  $\beta$ , the association of temperature and mortality, and its SE on each simulated dataset. Results were summarized across the 10,000 replications. Performance measures of primary interest included bias and variance. Specifically, we reported relative average bias, defined



Figure 2. Relative average bias for four methods under data-generating Scenario 1 (A) and Scenario 2 (B), with panels for each version of the daily minimum temperature variable. cts, continuous; numbers are percentile cutoffs for increasingly rare exposures.



Figure 3. Relative efficiency for four methods under data-generating Scenario 1 (A) and Scenario 2 (B), with panels for each version of the daily minimum temperature variable. cts, continuous; numbers are percentile cutoffs for increasingly rare exposures.

as the average across all 10,000 datasets of:  $(\beta - \beta)/\beta$ . We reported relative efficiency, consistent with prior literature,<sup>9</sup> defined as the ratio of the mean variances of  $\beta$  from each method relative to the mean-variance of  $\beta$  from an arbitrary reference method, TS<sub>8</sub> (i.e., SE<sup>2</sup><sub>TS8</sub>/SE<sup>2</sup>CC<sub>CP</sub>). Computation was done on the University of Southern California's high-performance computing cluster at the Center for Advanced Research Computing.

## Results

#### Simulation study

As expected, bias depended on the data-generating scenario. Under Scenario 1 (Figure 2A), the TS<sub>8</sub> model produced estimates with negligible bias (relative average bias across temperature variables: -5.6e-05% to -3.8e-03%) while the TS<sub>4</sub> model estimates using fewer df had positive bias (1.6% to 21%) and the CC<sub>CP</sub> estimates had negative bias (-16 to -3.8%). Conversely, under Scenario 2 (Figure 2B), the CC<sub>CP</sub> model produced the least

biased estimates, while the TS model estimates tended to have positive bias.

Regardless of the data-generating mechanism,  $CC_{CP}$  estimates had larger uncertainty than TS methods. This is reflected in the lower relative efficiency for  $CC_{CP}$  compared with TS models, with declines in relative efficiency as the temperature variable  $x_t$  becomes more rare (Figure 3). The more parsimonious TS<sub>4</sub> model had the highest relative efficiency across all versions of  $x_t$  (102%–138%), but approached the efficiency of TS<sub>8</sub> as  $x_t$ became more rare. The  $CC_{CP}$  model had worse relative efficiency across all  $x_t$  (80%–91%), worsening as  $x_t$  became more rare. While both  $CC_{SB1}$  and  $CC_{SB2}$  had similar declines in efficiency as  $x_t$  became more rare, relative efficiency for  $CC_{SB1}$  was worse (70%–80%) than CC or  $CC_{SB2}$  (85%–102%) (Table S3; http:// links.lww.com/EE/A328). Comparing statistical power for TS<sub>8</sub> under Scenario 1 and  $CC_{CP}$  under Scenario 2, both methods have >99% power using continuous, and 80th, 90th, and 95th percentile cutoffs for  $x_t$ . However, TS<sub>8</sub> has greater power than  $CC_{CP}$  at the higher cutoffs for  $x_t$ , (i.e., 97.5th percentile: 93% vs. 88%; 99th percentile: 63% vs. 54%).

Table 2.							
Dropping out of data in case-crossover method							
Temperature	All-	Unique strata	Unique strata with				
voriable	dootho n	(0/ of continuous)	ovtromo tomporatur				

variable	deaths, n	(% of continuous)	extreme temperature
Continuous	367,712	504	N/A <sup>a</sup>
80th	116,100	170 (33.7 %)	126
90th	83,653	122 (24.2 %)	63
95th	58,149	83 (16.4 %)	24
97.5th	33,582	47 (9.3 %)	7
99th	14,916	21 (4.2%)	1

Number of unique strata (referent windows including case day plus 3 or 4 referent days) in timestratified CC analyses using daily minimum temperature (Tmin), defined as continuous Tmin or a percentile-based dichotomization ( $x_t$ ), in LA County, CA from 1 January 2014 to 31 December 2019. <sup>a</sup>Not applicable under continuous temperature with no designated extreme temperature days.

In Scenarios 3 and 4, the magnitude of the relative average bias was smaller, but all trends in both relative average bias (Figure S1; http://links.lww.com/EE/A328) and relative efficiency (Figure S2; http://links.lww.com/EE/A328) were identical to the trends in Figures 2 and 3, respectively. Finally, our results for  $CC_{CP}$  were identical whether using the conditional logistic regression approach using weights for mortality counts ( $CC_{wtd}$ ) or using the traditional approach ( $CC_{ex}$ ; Table S3; http://links.lww.com/EE/A328). Comprehensive numeric simulation study results are provided in Table S2; http://links.lww.com/EE/A328.

#### Los Angeles County data analysis

There were 367,712 all-cause deaths in LA County from 1 January 2014 to 31 December 2019, and the average daily Tmin was 11.39 °C (standard deviation: 6.16 °C). When applying both TS<sub>8</sub> and CC<sub>CP</sub> methods to data in LA County we found that a 1 °C higher daily Tmin was associated with a 0.44% (95% confidence interval [CI]: 0.31%, 0.56%) increase in mortality according to TS<sub>8</sub> and a 0.26% (95% CI: 0.13%, 0.39%) increase in mortality according to CC<sub>CP</sub> (Table 1). Using the 99th percentile cutoff for dichotomization, extremely hot days were associated with a 6.8% (95% CI: 3.3%, 10.5%) increase in mortality according to TS and a 5.6% (95% CI: 1.9%, 9.8%) increase in mortality according to TS and a state than TS<sub>8</sub> estimates.

Insights into the reduced efficiency of CC methods may be gained by noting that for a dichotomous  $x_t$ , the only informative strata in a conditional regression are those containing at least one, but not all, days with extreme exposure. In LA County, for the 99th percentile of daily Tmin, only 4.2% of strata (21/504) had at least one extremely hot day (Table 2). Additionally, multiple extreme temperature days were frequently observed within a given strata (i.e., >1 of the case or control days had an extreme temperature), which would violate the transience of exposure assumption if these are from the same heatwave event. For example, using extreme heat at the 90th percentile, we observed nine heat waves from 2014 to 2019 that were >7 days in duration, including one that stretched 24 days.

## Discussion

We compared the performance of TS and CC methods in estimating the effects on mortality of extreme daily Tmin days in a simulation study based on data from LA County. CC was less statistically efficient (i.e., larger uncertainty of the estimated association) compared with TS, especially as the extreme exposures considered were increasingly rare. CC with rare exposures uses only a small fraction of the total dataset since strata not containing the rare exposure are dropped out in conditional regression models. We found that both methods yielded biased estimates when inadequately controlling for time trends, as has been previously reported.<sup>4,7,8,21</sup> When analyzing observed data in LA County, both methods estimated qualitatively similar associations, with larger association estimates and smaller SE for the TS method as compared to the CC method, in line with our simulation study results under both assumed data-generating scenarios.

The efficiency of TS and CC methods has been studied previously in air pollution epidemiology settings.<sup>15,21,22</sup> Bateson and Schwartz<sup>9</sup> conducted a simulation study for daily TS data, which included a comparison of the efficiency of TS and CC methods for both a continuous and binary exposure variable with a prevalence of 25% using versions of these methods popular at the time. Across various data-generating scenarios with different assumptions about seasonal and long-term time trends, they found that CC estimates had larger variance than Poisson regression, with relative efficiencies (variance of Poisson/variance of CC) ranging from 0.65 to 0.69, with no difference found between continuous and binary exposures. The CC sampling design used in their work was the 1-week SB design, similar to  $\mathrm{CC}_{\mathrm{SB1}}$  in our study.^{21,22} Our simulations using the more modern time-stratified CC version found relative efficiencies in the same direction but less extreme, likely because the time-stratified referent sampling scheme includes a larger number of control days (three or four) versus the two in the  $CC_{SB1}$  design, providing improved efficiency. Our simulations using  $CC_{SB1}$  had relative efficiencies closer in magnitude to those reported in Bateson and Schwartz. Our simulations provide new evidence that when a binary exposure variable is increasingly rare, this impacts the efficiency of CC models relative to TS models. Time-stratified CC and TS methods with a regression spline for time were also compared for estimating the association of particulate matter air pollution with mortality in a simulation by Lu et al.8 They found that a quasi-Poisson TS model with greater df (12 df/year) had approximately the same efficiency as the time-stratified CC model, but a TS model with smaller df (4 df/year) yielded greater efficiency than either other method. Our results corroborate their findings in the context of rare, dichotomized environmental exposures and show how these findings can be generalized to other types of exposures. Specifically, our results show that relative efficiencies of all TS models with differing df converge to a similar value as  $x_t$  becomes increasingly rare, while the relative efficiency of the time-stratified CC model decreases. It follows that the number of df included in the regression spline in a TS model does not account for all of the differences in uncertainty between TS and CC models when assessing the association of rare exposures. Hence, the efficiency of these methods and the choice of study design should take into consideration the rarity of exposure. Of note, the choice of df for TS has important implications for bias, regardless of the rarity of the exposure, but appears to have less impact on efficiency for the most rare exposures.

Either TS or CC methods can yield biased estimates when inadequately controlling for time trends<sup>4,7,8</sup> and our results corroborate these findings. CC designs are said to control for time-varying confounders by design by only making withinparticipant comparisons around a short frame of time. However, residual seasonal confounding and day-to-day variation in confounding variables need to be accounted for separately.<sup>21</sup> A TS analysis must carefully control for seasonality and day-to-day variation (i.e., using BIC to select between optimal df in a natural cubic spline of time) trying to avoid overfitting or underfitting time trends.<sup>6</sup> Worries about residual seasonal confounding of both methods have led some to limit their analyses to only a few months out of the year, as in Sun et al. and Gasparrini et al.<sup>23,24</sup>

Our study has several strengths. First, we conducted a realistic simulation study based on observed data in LA County, a location with a large population that experience days with extreme heat. Second, in our simulations, we fixed the association between the exposure, confounders, and the outcome to target only the increasing rarity of exposure days or both increasingly rare and extreme associations. Third, the data-generating scenarios used represented time trends similar to assumptions underlying TS (Scenarios 1 and 3) or time-stratified CC approaches (Scenarios 2 and 4). Fourth, we used the conditional Poisson implementation of CC that requires no complicated data management, is computationally efficient, and is equivalent to traditional conditional logistic regression implementations. Additional advantages of this approach are that it can be extended to account for overdispersed (quasi-Poisson), autocorrelated count data, and subject-level covariates and exposures.<sup>16,20</sup> Future work could extend the comparison to CC using individualized exposures and multisite or participant-level TS. Finally, while we focused on extreme heat as our exemplar exposure, our methodological findings apply to other extreme weather exposures.

Our study has several limitations and raises issues requiring further research. First, our study only addresses day-to-day variation in extreme heat exposure. Heatwaves often occur over consecutive days and may violate assumptions regarding the transience of exposure for both TS and CC methods. This problem will be exacerbated by the increasing intensity and duration of heat waves expected under the changing climate.<sup>25</sup> More sophisticated definitions of extreme heat events, such as consecutive extremely hot days, or making use of rolling averages should be considered, as well as alternatives to TS and CC methods that focus on exposure event(s), such as studies of excess mortality from specific events.26 Our study used percentile thresholds to define extreme heat based on the concept of acclimatization. Thresholds for extreme heat based on absolute values (e.g., days exceeding 43 °C daily maximum) could also be explored, but the general methodological findings in our simulation study will still apply. We considered dichotomized temperature variables for simplicity, but the performance of nonlinear effect estimation for continuous temperature variables could also be studied. We used the Poisson distribution both to generate mortality data and Poisson regression to estimate associations (similar to prior literature<sup>9</sup>) rather than quasi-Poisson regression, which is a generalization that includes an overdispersion parameter allowing for the variance to be unequal to the mean and can increase estimation uncertainty. We compared our results to those of Lu et al., who compared quasi-Poisson TS to CC and they found the standard deviation of the quasi-Poisson TS was slightly larger than CC for a greater df (12 df/year), and considerably smaller for a smaller df (4 df/year). Another simplification that we made was to only assess the effect of extremely high temperatures on mortality, not extremely low temperatures. This was informed by studies that attribute small impacts on mortality for cold temperatures in LA County.<sup>27</sup> Finally, the impacts of heat on mortality vary by factors such as participant age, air conditioning access/use, and time-varying covariates such as humidity, which were not considered since our study focused on the main effect of heat exposure and our synthetic mortality datasets assumed no effect modification.

In conclusion, our study provides insight into two prominent statistical methodology choices for use in future studies of extreme temperatures, one of the most commonly studied climate-related hazards.<sup>11,27</sup> However, our findings also extend to other climate change and health studies relating extreme and rare exposures to large, temporally-resolved administrative health datasets. In practice, CC has been used to estimate the effect of rare weather and air pollution exposures.<sup>3,12,23,28</sup> As the climate continues to change, extreme weather events will become more frequent and intense.<sup>1</sup> Researchers studying these events should consider both TS and CC approaches to carefully decide which method is most applicable to their data and most appropriate for their intended research question. TS approaches can provide more precise estimates based on the rare extreme exposure data available today to project public health impacts of the future increasingly common extreme exposures, assuming appropriate control for time trends.

## **Conflicts of interest statement**

The authors declare that they have no conflicts of interest with regard to the content of this report.

#### References

- 1 Romanello M, Napoli CD, Green C, et al. The 2023 report of the Lancet Countdown on health and climate change: the imperative for a health-centred response in a world facing irreversible harms. *Lancet*. 2023;402:2346–2394.
- 2 Fouillet A, Rey G, Laurent F, et al. Excess mortality related to the August 2003 heat wave in France. *Int Arch Occup Environ Health*. 2006;80:16–24.
- 3 Saucy A, Ragettli MS, Vienneau D, et al. The role of extreme temperature in cause-specific acute cardiovascular mortality in Switzerland: a case-crossover study. *Sci Total Environ.* 2021;790:147958.
- 4 Bhaskaran K, Gasparrini A, Hajat S, Smeeth L, Armstrong B. Time series regression studies in environmental epidemiology. *Int J Epidemiol.* 2013;42:1187–1195.
- 5 Perperoglou A, Sauerbrei W, Abrahamowicz M, Schmid M. A review of spline function procedures in R. BMC Med Res Methodol. 2019;19:46.
- 6 Peng RD, Dominici F, Louis TA. Model choice in time series studies of air pollution and mortality. J R Stat Soc Ser A Stat Soc. 2006;169:179–203.
- 7 Janes H, Sheppard L, Lumley T. Case-crossover analyses of air pollution exposure data: referent selection strategies and their implications for bias. *Epidemiology*. 2005;16:717–726.
- 8 Lu Y, Symons JM, Geyh AS, Zeger SL. An approach to checking case-crossover analyses based on equivalence with time-series methods. *Epidemiology*. 2008;19:169–175.
- 9 Bateson TF, Schwartz J. Control for seasonal variation and time trend in case-crossover studies of acute effects of environmental exposures. *Epidemiology*. 1999;10:539–544.
- 10 Carracedo-Martínez E, Taracido M, Tobias A, Saez M, Figueiras A. Case-crossover analysis of air pollution health effects: a systematic review of methodology and application. *Environ Health Perspect*. 2010;118:1173–1182.
- 11 Harper SL, Cunsolo A, Babujee A, et al. Trends and gaps in climate change and health research in North America. *Environ Res.* 2021;199:111205.
- 12 Rahman MM, McConnell R, Schlaerth H, et al. The effects of coexposure to extremes of heat and particulate air pollution on mortality in California: implications for climate change. Am J Respir Crit Care Med. 2022;206:1117–1127.
- 13 Abatzoglou JT. Development of gridded surface meteorological data for ecological applications and modelling. *Int J Climatol.* 2013;33:121–131.
- 14 Medina-Ramón M, Zanobetti A, Cavanagh DP, Schwartz J. Extreme temperatures and mortality: assessing effect modification by personal characteristics and specific cause of death in a multi-city case-only analysis. *Environ Health Perspect*. 2006;114:1331–1336.
- 15 Fung KY, Krewski D, Chen Y, Burnett R, Cakmak S. Comparison of time series and case-crossover analyses of air pollution and hospital admission data. *Int J Epidemiol.* 2003;32:1064–1070.
- 16 Wu Y, Li S, Guo Y. Space-time-stratified case-crossover design in environmental epidemiology study. *Health Data Sci.* 2021;2021:9870798.
- 17 Friedrich S, Friede T. On the role of benchmarking data sets and simulations in method comparison studies. *Biom J.* 2024;66:e2200212.
- 18 Dionisio KL, Chang HH, Baxter LK. A simulation study to quantify the impacts of exposure measurement error on air pollution health risk estimates in copollutant time-series models. *Environ Health*. 2016;15:114.
- 19 Boulesteix A-L, Groenwold RH, Abrahamowicz M, et al; STRATOS Simulation Panel. Introduction to statistical simulations in health research. *BMJ Open*. 2020;10:e039921.
- 20 Lu Y, Zeger SL. On the equivalence of case-crossover and time series methods in environmental epidemiology. *Biostatistics*. 2007;8:337–344.

- 21 Whitaker HJ, Hocine MN, Farrington CP. On case-crossover methods for environmental time series data. *Environmetrics*. 2006;18:157–171.
- 22 Bateson TF, Schwartz J. Selection bias and confounding in casecrossover analyses of environmental time-series data. *Epidemiology*. 2001;12:654–661.
- 23 Sun S, Weinberger KR, Nori-Sarma A, et al. Ambient heat and risks of emergency department visits among adults in the United States: time stratified case crossover study. *BMJ*. 2021;375:e065653.
- 24 Gasparrini A, Guo Y, Hashizume M, et al. Changes in susceptibility to heat during the summer: a multicountry analysis. Am J Epidemiol. 2016;183:1027–1036.
- 25 Calvin K, Dasgupta D, Krinner G, et al. IPCC, 2023: Climate Change 2023: Synthesis Report. Contribution of Working Groups

I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [core writing team, H. Lee and J. Romero (eds.)]. Intergovernmental Panel on Climate Change (IPCC); 2023. doi:10.59327/ipcc/ar6-9789291691647.

- 26 Hoshiko S, English P, Smith D, Trent R. A simple method for estimating excess mortality due to heat waves, as applied to the 2006 California heat wave. *Int J Public Health*. 2010;55:133–137.
- 27 Weinberger KR, Haykin L, Eliot MN, Schwartz JD, Gasparrini A, Wellenius GA. Projected temperature-related deaths in ten large U.S. metropolitan areas under different climate change scenarios. *Environ Int.* 2017;107:196–204.
- 28 Wilson LA, Morgan GG, Hanigan IC, et al. The impact of heat on mortality and morbidity in the Greater Metropolitan Sydney Region: a case crossover analysis. *Environ Health*. 2013;12:98.