Supplementary Material

Statistical model

Here we provide more details on the model described by (1) and (2).

We assume $\alpha(t)$ is smooth enough that it can be represented by a smoothing cubic spline with one knot every seven years. If we have less than seven years of data, then $\alpha(t)$ is simply a linear function of time. To represent the seasonal trend we assumed a *harmonic model*:

$$s(t) = \sum_{k=1}^{K} \{a_k \cos(2\pi kt/365) + b_k \sin(2\pi kt/365)\},$$

where the parameters $\boldsymbol{a} = [a_1, \dots, a_K]$ and $\boldsymbol{b} = [b_1, \dots, b_K]$ are estimated from data. Moreover, we model the weekday-specific effects using seven indicator variables and seven constrained parameters:

$$w(t) = \sum_{d=1}^{7} w_d X_d(t), \; \sum_{d=1}^{7} w_d = 0$$

with $X_d(t) = 1$ if day t is day of the week d and 0 otherwise. Exploratory data analysis demonstrated that this source of variability is pronounced in younger age groups that have higher death rates during the weekends. We note that if this source of variability is unaccounted for, the estimation of the parameters defining the error structure are affected.

We further assumed the vector $\boldsymbol{\varepsilon} = [\varepsilon_1, \dots, \varepsilon_T]^\top$ follows a truncated multivariate distribution with mean 1 (no change). To account for natural correlated variability we assumed $\boldsymbol{\varepsilon}$ had variance-covariance matrix, denoted with $\boldsymbol{\Sigma}$, determined by an auto-regressive (AR) process of order p:

$$\text{var}(\boldsymbol{\varepsilon}) \equiv \boldsymbol{\Sigma} = \sigma^2 \begin{pmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{T-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{T-2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho_{T-1} & \rho_{T-2} & \rho_{T-3} & \dots & 1 \end{pmatrix}.$$

with $\rho_1, \dots \rho_{T-1}$ determined by the parameters of the AR(p) process. To assure $\mathbb{E}(Y_t \mid \varepsilon_t) > 0$ we assumed that $\Pr(\varepsilon_t > 0) = 1$. Note that the percent change in death rate due to natural variation, not accounted for seasonality and secular trends, observed in practice is exclusively between 20% smaller and 20% larger than mortality ($0.8 < \varepsilon_t < 1.2$), which consistent with this assumption.

Estimating event effects

Here we describe the three-step approach that we use to estimate $f(t)$ and standard error for these estimates. The general idea is to first estimate μ_t and $\boldsymbol{\Sigma}$ during periods with control periods, and then estimate the most interest component: $f(t)$.

The first step is to estimate $\alpha(t), s(t)$, and $w(t)$. To do this, we select a control period I_{control} for which we know there were no natural disasters nor outbreaks and can assume $f(t) = 0$. Natural disasters and outbreaks are rare, hence, it should be possible to find such periods for most datasets. With this assumption in place model (1) reduces to:

$$Y_t \mid \varepsilon_t \propto \text{Poisson}(\mu_t \varepsilon_t), t \in I_{\text{control}}$$

Note that if we have $N \equiv \lfloor T/365 \rfloor$ years of daily data, given the choices described in the Tuning parameters section of the eAppendix, we are only fitting N/7 + 7 + 4 + 1 parameters to T data points. As an example, for seven years of data this translates into 2,556 data points and 13 parameters. As a result, we can obtain highly precise estimates $\hat{\mu}_t$ even in the presence of the extra dispersion introduced by ε_t. We therefore fit a quasi-Poisson Generalized Linear Model (GLM) to regions in I_{control} and estimate the expected value for $t \notin I_{\text{control}}$ using the newly learned parameters. The quasi-Poisson assumption permits us to model the extra variability introduced by ε_t.

In the second step we use the control region to estimate the variance-covariance matrix $\mathbf{\Sigma}$ as described in detail in the eAppendix. With this estimate in place we then use an iterative generalized least square procedure to estimate $f(t)$ and its standard error. We use the Central Limit Theorem approximation to assume $\hat{f}(t)$ follows a normal distribution. As explained in the eAppendix, this standard error includes the variability introduced by the uncertainty in the estimate of the expected mortality rate $\hat{\mu}_t$.

Estimating standard errors

The first step is to estimate the variance-covaiance matrix $\mathbf{\Sigma}$. We use data in the control period to do this. First, let r_t be the observed percent change from expected mortality:

$$r_t = \frac{Y_t - \hat{\mu}_t}{\hat{\mu}_t}$$

To propagate the uncertainty in the estimation process in the first step, we employ a first-order Taylor approximation of r_t. Note that $\mathbb{E}[\hat{\mu}_t] = \mu_t$ and $\mathbb{E}[Y_t] = \mu_t$, where the latter equality holds by the law of total expectation. Then, we can approximate r_t around (μ_t, μ_t) and find that:

$$\mathbb{E}\left(r_t\right) \approx 0 \text{ and } \text{Var}\left(r_t\right) \approx \sigma^2 + 1/\hat{\mu}_t + \text{Var}\left(\log \hat{\mu}_t\right)$$

where we used the law of total variance to find $\text{Var}(Y_t)$. Intuitively, the terms in $\text{Var}(r_t)$ represent the variance added by ε, the Poisson variability, and the uncertainty from the first step, respectively. The above implies that the following random variable:

$$Z_t = \frac{r_t}{\sqrt{\text{Var}\left(r_t\right)}}$$

has the same correlation structure as ε_t. We therefore use the Yule-Walker equations to estimate the AR process parameters from the observed Z_t within I_{control}. To estimate σ^2

we use:

$$\hat{\sigma}^2 = \max\left\{\frac{1}{T}\sum_{t=1}^{T}\left[\left(\frac{Y_t - \hat{\mu}_t}{\hat{\mu}_t}\right)^2 - \frac{1}{\hat{\mu}_t} - \text{Var}(\log\hat{\mu}_t)\right], 0\right\}$$

Using these we can form an estimate $\hat{\mathbf{\Sigma}}$.

In the third and final step we estimate $f(t)$. Denote $\mathbf{r} = (r_1 \dots, r_T)^\top$, $\mathbf{f} = (f(1), \dots, f(T))^\top = \mathbf{B}\boldsymbol{\theta}$, with $\mathbf{B}$ and $\boldsymbol{\theta}$ the design matrix and parameters, respectively, that define the natural cubic spline and $\mathbf{D}$ a diagonal matrix with entries:

$$d_{t,t} = \sqrt{[1 + f(t)]^2 + \frac{1 + f(t)}{\hat{\mu}_t\hat{\sigma}^2} + \frac{1}{\hat{\sigma}^2}[1 + f(t)]^2\,\text{Var}(\log\hat{\mu}_t)}$$

Then, we can use the fact that

$$\mathbb{E}(\mathbf{r}) = \mathbf{B}\boldsymbol{\theta} \text{ and } \text{Var}(\mathbf{r}) = \mathbf{D\Sigma D}$$

to obtain an unbiased estimate of $\mathbf{f}$ using generalized least squares:

$$\hat{\boldsymbol{\theta}} = (\mathbf{B}^\top(\mathbf{D\Sigma D})^{-1}\mathbf{B})^{-1}\mathbf{B}^\top(\mathbf{D\Sigma D})^{-1}\mathbf{r}$$

along with a variance estimate:

$$\text{Var}(\boldsymbol{\theta}) = (\mathbf{B}^\top(\mathbf{D\Sigma D})^{-1}\mathbf{B})^{-1}$$

Since $\mathbf{D\Sigma D}$ depends on the estimand $\mathbf{f}$, we use an iterative procedure in which we plug-in the current estimate to compute the variance.

Excess mortality estimate

To estiamte the standard error for $\hat{\Delta}_{[t_0,t_1]}$, let $\boldsymbol{B}$ and $\boldsymbol{\theta}$ be as they were defined above. Then, note that we can conveniently represent excess deaths at time t as $\mu_t \times f(t)$ and define excess deaths for any time period $[t_0, t_1]$ by just adding these up:

$$\Delta_{[t_0,t_1]} = \sum_{t=t_0}^{t_1}\mu_t \times f(t) = (\mu_{t_0}, \dots, \mu_{t_1})\,\boldsymbol{B}_{[t_0:t_1,]}\boldsymbol{\theta}$$

where $\boldsymbol{B}_{[t_0:t_1,]}$ represents rows t_0,. . . ,t_1 of matrix $\boldsymbol{B}$. We therefore estimate cumulative excess deaths for any interval $I = [t_0, t_1]$ by adding the excess deaths for each day in I:

$$\hat{\Delta}_{[t_0,t_1]} = \sum_{t=t_0}^{t_1}\hat{\mu}_t \times \hat{f}(t) = (\hat{\mu}_{t_0}, \dots, \hat{\mu}_{t_1})\,\boldsymbol{B}_{[t_0:t_1,]}\hat{\boldsymbol{\theta}}$$

where $\hat{\boldsymbol{\theta}}$ is the Maximum Likelihood estimate of $\boldsymbol{\theta}$. We construct a 95% confidence interval using the CLT approximation to assume the sum is approximately normally distributed with variance given by:

$$\text{Var}\left[\sum_{t=t_0}^{t_1} \hat{\mu}_t \times \hat{f}(t)\right] = (\hat{\mu}_{t_0}, \ldots, \hat{\mu}_{t_1})\boldsymbol{B}_{[t_0:t_1,]} \left[\boldsymbol{B}^\top (\mathbf{D}\boldsymbol{\Sigma}\mathbf{D})^{-1}\boldsymbol{B}\right]^{-1} \boldsymbol{B}^\top_{[t_0:t_1,]}(\hat{\mu}_{t_0}, \ldots, \hat{\mu}_{t_1})^\top.$$

Note that our excess deaths estimate $\hat{\mu}_t \times \hat{f}(t)$ for time t is a smooth version of the estimate based on single time point: $Y_t - \hat{\mu}_t$. In fact, our estimate converges to $Y_t - \hat{\mu}_t$ when the number of knots that define f is equal to the number of observations. The corresponding argument in our software package is the `knots_per_year` argument in the `excess_model` function.

Mortality data stratified by demographic indicators

When mortality data is stratified by demographic indicators, such as age and sex, we can estimate event effect for each group. Note that we can then use this to estimate adjusted marginal effect. Specifically, if $\hat{f}_k(t)$ is the estimated effect at time t in group k we can define the adjusted overall-effect as event effect as a weighted sum of the group-specific effects:

$$\hat{f}(t) = \sum_{k=1}^{K} \pi_k(t) \times \hat{f}_k(t), \text{ where } \pi_k(t) \equiv \frac{\hat{\mu}_{k,t}}{\sum_{i=1}^{K} \hat{\mu}_{i,t}}$$

and $\hat{\mu}_{k,t}$ is the expected value of group k at time t. As before, we can approximate $\text{Var}\left(\hat{f}(t)\right)$ with a first-order Taylor approximation around $(f_1(t), \ldots, f_K(t), \mu_{1,t}, \ldots, \mu_{K,t})$ and find that:

$$\text{Var}\left(\hat{f}(t)\right) \approx \sum_{k=1}^{K} \left[\pi_{k,1}(t)\text{Var}\left(\hat{f}_k(t)\right) + \pi_{k,2}(t)\text{Var}\left(\hat{\mu}_{k,t}\right)\right]$$

, where

$$\pi_{k,1} \equiv \left(\frac{\hat{\mu}_{k,t}}{\sum_{i=1}^{K} \hat{\mu}_{i,t}}\right)^2 \text{ and } \pi_{k,2} \equiv \left(\frac{\hat{f}_k(t)}{\sum_{i=1}^{K} \hat{\mu}_{i,t}} - \frac{\sum_{k=1}^{K} \hat{\mu}_{k,t} \times \hat{f}_k(t)}{\left(\sum_{i=1}^{K} \hat{\mu}_{i,t}\right)^2}\right)^2$$

Tuning parameters

Our approach included four fixed parameters that need to be defined before fitting the model to data. Here we motivate the default values we have set in the software implementation.

- We represent $\alpha(t)$ as a smoothing cubic spline with one knot every seven years. If we have less than seven years of data, then $\alpha(t)$ is simply a linear function of time. This choice is motivated by the observation that in many jurisdictions death rates have been slowly declining for the past decades.

- The seasonal trend is estimated by $s(t)$, a periodic function with K harmonics. By computing and plotting the daily average count for each day of the year, for the 35 years of the Puerto Rico dataset, we noted that the pattern was not exactly sinusoidal (K=1) but that adding one harmonic (K=2) captured the extra complexity for most jurisdictions. We recommend generating this plot to guide the choice of K and we provide software in our package to do so.

- We can use our approach in an exploratory mode to search for periods in which $f(t) \neq 0$. To do this, we use a flexible spline with 12 knots per year. This allows sharp increases to be captured. However, once we discover a period of increased death rates, we generate exploratory plots to determine if a smoother estimate is more appropriate. We used a spline with six knots per year for all hurricane effects and a discontinuity on landfall day.

- We set the degree of the AR process at seven. This choice was made by examining residual auto-correlation plots and assessments based on comparing observed and predicted standard errors for excess mortality estimates generated from periods for which we expected $f(t) = 0$. The software implementation permits the user to change this parameter or use AIC to determine the degree, as implemented in the R function `ar`.

Population size estimates

We used yearly population estimates from the US Census, which correspond to the population size on July 1, to estimate daily population counts, N_t, via linear interpolation. For days past the last value from the US Census, we assumed the population was constant and equal to the last day for which we had data. Yearly population estimates for Puerto Rico were obtained from the Puerto Rico Institute of Statistics (PRIS). Similarly, we estimated N_t via linear interpolation of the observed values. However, to account for population displacement after hurricane María [27, 28], we used estimates based on mobile phone records provided by Teralytics —a technology company that partners with telecommunication operators worldwide to assess human mobility[29]. Specifically, Teralytics provided daily population proportion estimates relative to a confidential baseline from May 2017, to April 2018. We multiplied these proportions by the 2017 mid-year population value from PRIS and generated smooth estimates with local regression. Finally, from the United Nations, we obtained age-specific population proportions in five year intervals for Puerto Rico from 1950 to 2020. Since we only have mortality data for Puerto Rico dating back to 1985, we generated daily demographic proportion estimates via linear interpolation of the five year interval values starting in 1985. Then, we computed age-specific population estimates by multiplying the demographic proportions times the aforementioned smooth population estimates.

Simulation studies

To assess our procedure we conducted a Monte Carlo simulations. We designed simulation studies to mimic three scenarios 1) a natural disaster, 2) an infectious disease epidemic, and 3) a typical period with no events. The event effect, $f(t)$, for the natural disaster scenario was defined by a strong direct effect followed by a slow decaying indirect effect, as seen after natural disasters. To achieve this, we fit model (2) to mortality data from Puerto Rico for all deceased 75 and older to periods with no known events. This provides estimates for $\hat{\alpha}(t)$, $\hat{s}(t)$, and $\hat{w}(t)$ (eFigure 13) which we use to define μ_t for our simulations. We then fit model (1) to the period of July 20, 2017, to April 20, 2018, which includes Hurricane María, and used the resulting estimate $\hat{f}(t)$ as the true $f(t)$ for the first simulation study. To generate realistic effects $f(t)$ for the infectious disease epidemic scenario, we fit model (1) from May 14, 2014, to February 14, 2015, which includes the Chikungunya outbreak in Puerto Rico, and use the resulting $\hat{f}(t)$ as the true $f(t)$. Finally, we let $f(t) = 0$ for the entire period for the typical scenario simulation.

For natural variability we simulated serially correlated random variables $\{\varepsilon_1, \dots, \varepsilon_T\}$ centered at one with an AR process of degree two and coefficients obtained using the Yule-Walker equations (see eAppendix for details). We set the standard deviation of ε_t to 0.05 and used data from January 1, 2002 to December 31, 2013, a contiguous time period in which no known events occurred, to estimate the coefficients. With these parameters in place we then generated $B = 100,000$ simulated datasets following model (1):

$$Y_t^{(b)} \sim \text{Poisson}\left(\mu_t[1 + f(t)]\varepsilon_t^{(b)}\right) \text{ for } t = 1, \dots, T$$

for $b = 1, \dots, B$. Note that $Y_t^{(b)}$ refers to the simulated number of deaths at time t for simulation b. In eFigure 1, the solid-red curves represent the true curve or true standard error, whereas the dashed-black curves correspond to our estimates. We find that our method consistently estimates the true curve precisely under all three scenarios (eFigures 1A-C). Our estimated standard error also estimates the true standard error precisely (eFigures 1D-F).

Small death counts

The smoothing approach we employ improves power over methods based on measurements from a single day, particularly in situations where the signal to noise ratio is low. This implies that our method can be useful even when the death counts per time unit are small. We conducted a simulation study to determine the lower limit for deaths per day rates that our approach can handle. Specifically, we simulated data using the same approach we described above, but we normalized the counts such that the average deaths per day was $\gamma = \{0.05, 0.10, 0.50, 1.00\}$ across the simulated data. Specifically, we use:

$$Y_t^{(b)} \sim \text{Poisson}\left(\gamma \times \frac{\mu_t[1 + f(t)]}{\sum_{j=1}^{T} \mu_j[1 + f(j)]}\varepsilon_t^{(b)}\right) \text{ for } t = 1, \dots, T$$

The simulation demonstrated that for daily count rates above 0.10 our model performs well (eTable 1, eFigures 2 & 3). Finally, we note that for the demographics groups described in Section and used in Sections and , the lowest counts per day rate was 0.30.

Amount of training data

The first step of our approach is to estimate the expected counts μ_t for each time t using data in the control region. To assess the impact of the length of this control baseline period in our estimation procedure, we employed the same simulation scheme we described above, but we limited the length of the control region to 2, 4, 6, and 8 years. We found that results were practically equivalent for all training periods (eTable 4, eFigures 4 & 5)

False Discovery Rates and Power Analysis

To study the false discovery rate of our procedure we again repeated the simulation used in Section using the null model $f(t) = 0$. For each simulated dataset, we fit our model with 6 and 12 knots per-year. We also considered the saturated model that results in the estimate $\hat{f}(t) = (Y_t - \hat{\mu}_t)/\hat{\mu}_t$. For each model fit, we then searched for regions and recorded recorded all intervals for which $\hat{f}(t)$ was statistically different from 0 and the length of this interval. We then considered several period length requirements and reported the number of detected false events per year.

We repeated the simulation above but this time for a case in which $f(t) > 0$. Specifically, we defined f using the Tukey tri-weight function:

$$f(t) = aW\left(\frac{t - t_0}{b}\right) \text{ for } t = 1, \dots, T,$$

where t_0 is the day where the effect is highest, a is the peak effect, $2b$ is the length of the effect in days, and $W(u) = (1 - |u|^3)^3$, for $|u| \leq 1$. For this simulation we set $a = 0.20$ and $b = 45$, hence the period of indirect effect lasted 90 days. Similar to above, for each period length requirement, we recorded the proportion of years where one of our detected intervals had a non-null intersection with the interval for which $f(t) > 0$, specifically $[t_0 - b, t_0 + b]$. We reported the proportion of years for which the event of concern was detected.

We found that our method greatly improved sensitivity over the saturated model approach without much loss of specificity (eTables 2 & 3). If we require to see at least 5, 10, or 30 days above the threshold with a 12 knot smoother, then we see a false event every 2, 3, and 33 years, respectively, As described in the Methods section, we can also control specificity and sensitivity by increasing the confidence level.

Natural variability and correlated counts

To demonstrate the perilous effects of incorrectly assuming independence between counts, when in fact they are correlated, we compared our method to a Poisson and over-dispersed

Poisson model that assumes independent observations. We fit each of the three models to the Puerto Rico data for individuals 75 and over. Then, we randomly selected 100 intervals of sizes $L = 10$, 50, and 100 days from periods with no events, and computed the total number of deaths in each interval $S_l = \sum_{t \in L} Y_t$. We then used the fitted models to estimate the expected value $\hat{\mathbb{E}}(S_l)$ and standard error $\hat{\text{SE}}(S_l)$. If the expected value and the variance are estimated correctly, the Central Limit Theorem predicts that the statistic $Z_l = [S_l - \hat{\mathbb{E}}(S_l)]/\hat{\text{SE}}(S_l)$ should have an approximate Gaussian distribution with mean zero and variance one. We found that the Poisson and over-dispersed Poisson model underestimates the $\text{SE}(S_l)$ resulting in larger than expected values of $|Z_l|$.

Supplementary Tables

Counts per day	Estimate median bias	SE estimate median bias	SE estimate median RMSE
1.00	0.0004	0.0003	0.0082
0.50	0.0005	0.0013	0.0156
0.10	0.0030	0.0582	0.1073
0.05	0.0153	0.1277	0.2324

Supplementary Table 1: Assessment of the performance of our model when the rate of deaths per day is small. The first column shows the rate of deaths per day in each simulation. The second column shows the median, over time, of the absolute value of the bias of our estimate of the *event effect*. The third column shows the median, over time, of the absolute value of the bias of our estimate of the standard error of $\hat{f}$. The fourth column shows the median, over time, root mean squared error of our estimate of the standard error of $\hat{f}$.

knots	$\geq$ 1 day	$\geq$ 3 days	$\geq$ 5 days	$\geq$ 10 days	$\geq$ 1 month	$\geq$ 2 months
6	0.327	0.324	0.318	0.295	0.140	0.011
12	0.561	0.549	0.525	0.424	0.038	0.000
Saturated*	4.659	0.001	0.000	0.000	0.000	0.000

Supplementary Table 2: False discovery rates based on a simulation study where $f(t) = 0$, for all t. The results are shown as the rate of false events detected per year, out of the 100,000 simulated years. The first column shows the amount of smoothness. The second column shows results for all detected regions. Column 3 through 7 shows results for regions of length 3, 5, 10, 30, 60 or larger, respectively.

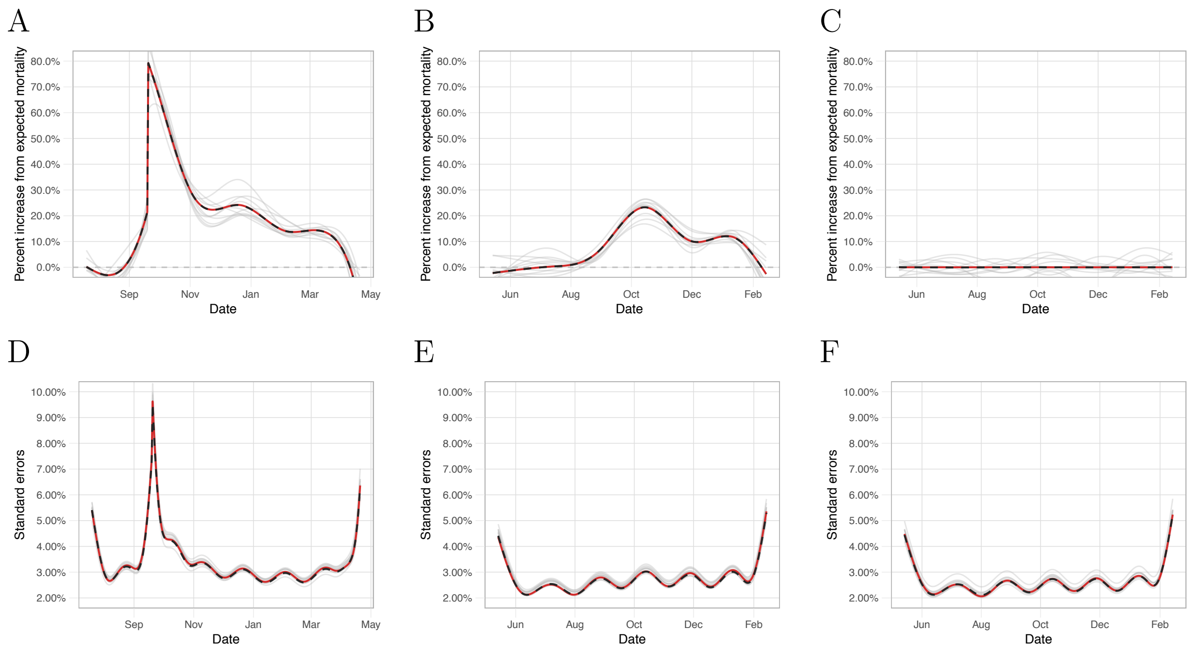
knots	$\geq$ 1 day	$\geq$ 3 days	$\geq$ 5 days	$\geq$ 10 days	$\geq$ 1 month	$\geq$ 2 months
6	1.000	1.000	1.000	1.000	1.000	0.999
12	1.000	1.000	1.000	1.000	1.000	0.762
Saturated*	1.000	0.119	0.003	0.000	0.000	0.000

Supplementary Table 3: Power analysis based on a simulation study where $f(t) > 0$ for an interval of 90 days. The results are shown as the rate of years in which we correctly detect the event of concern, out of the 100,000 simulated years. The first column shows the amount of smoothness. The second column shows results for all detected regions. Column 3 through 7 shows results for regions of length 3, 5, 10, 30, 60 or larger, respectively.

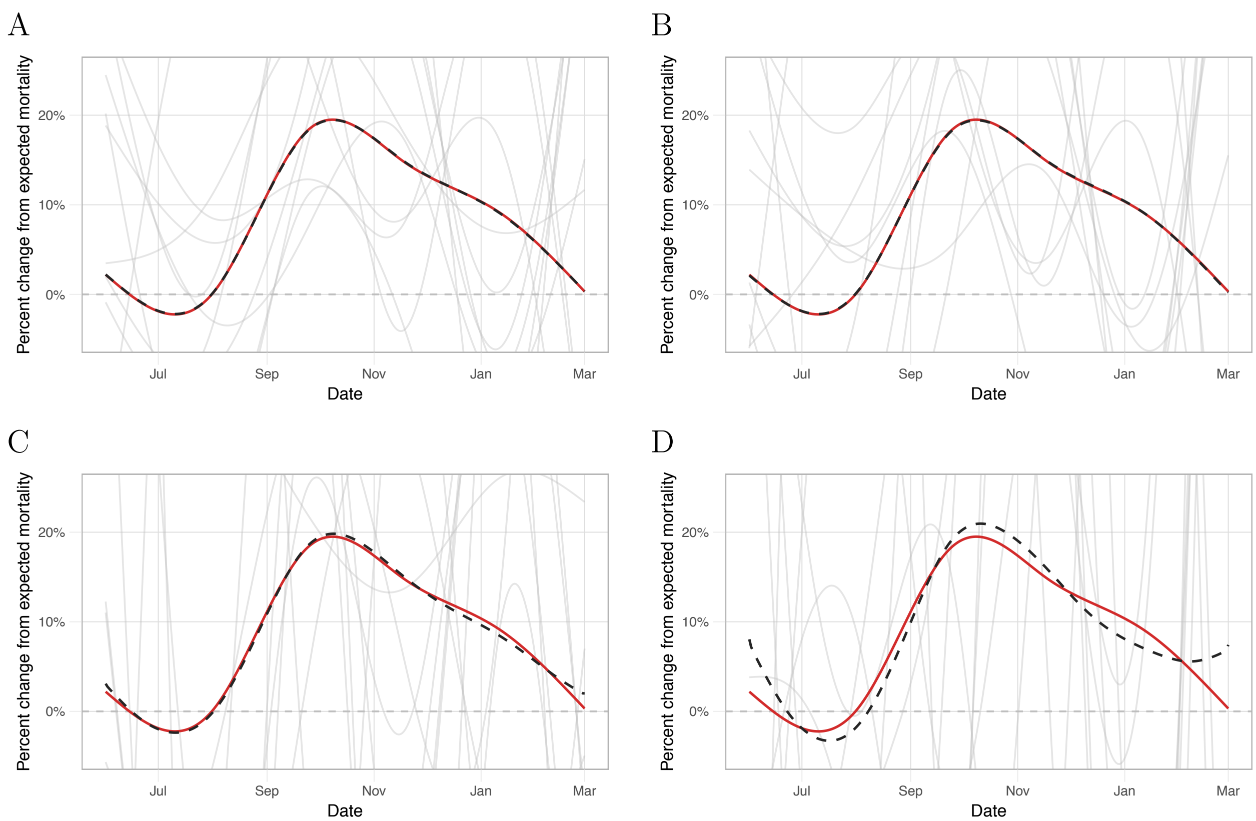
Length of the control region	Estimate median bias	SE estimate median bias	SE estimate median RMSE
8	0.0001	0.0035	0.0041
6	0.0001	0.0023	0.0032
4	0.0001	0.0031	0.0043
2	0.0001	0.0020	0.0040

Supplementary Table 4: Assessment of the effect of the length of the control period used to estimate the expected counts. The first column shows the rate of deaths per day in each simulation. The second column shows the median, over time, of the absolute value of the bias of our estimate of the *event effect*. The third column shows the median, over time, of the absolute value of the bias of our estimate of the standard error of $\hat{f}$. The fourth column shows the median, over time, root mean squared error of our estimate of the standard error of $\hat{f}$.

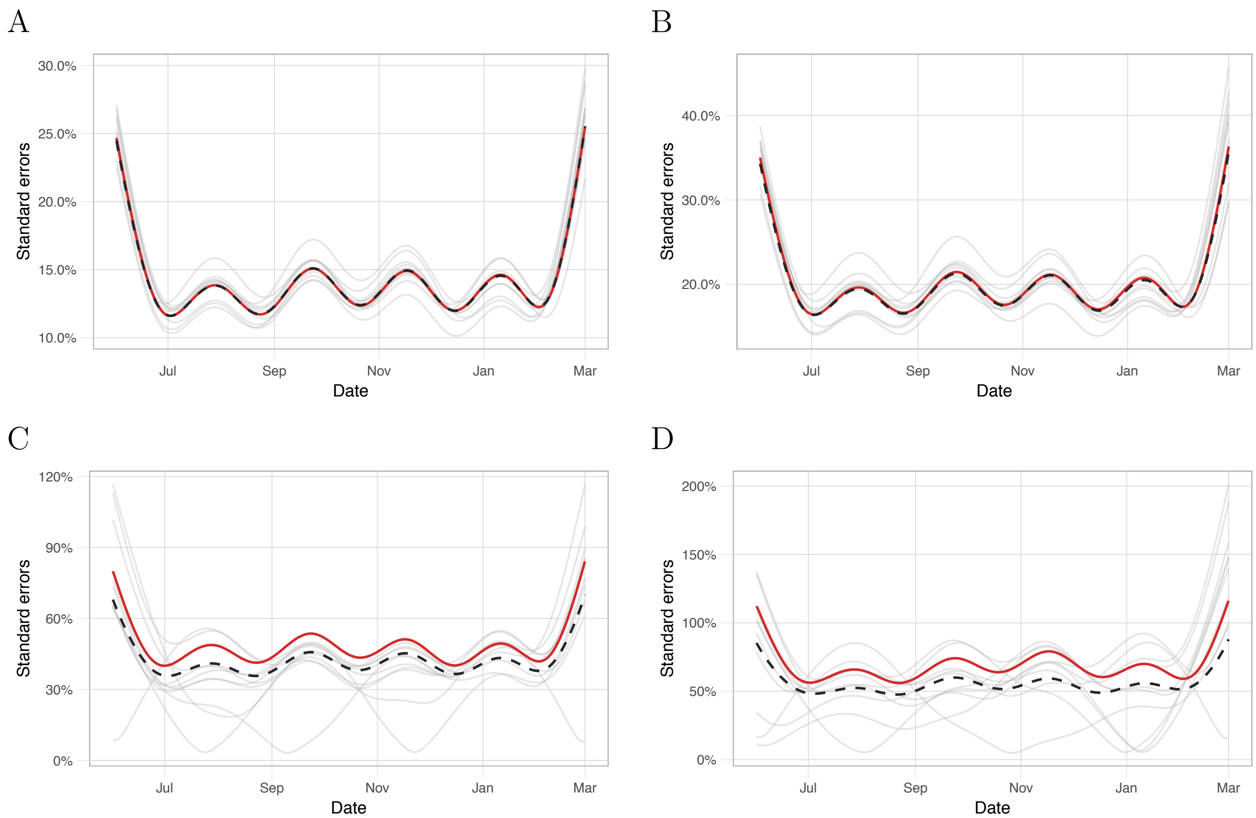
Supplementary Figures



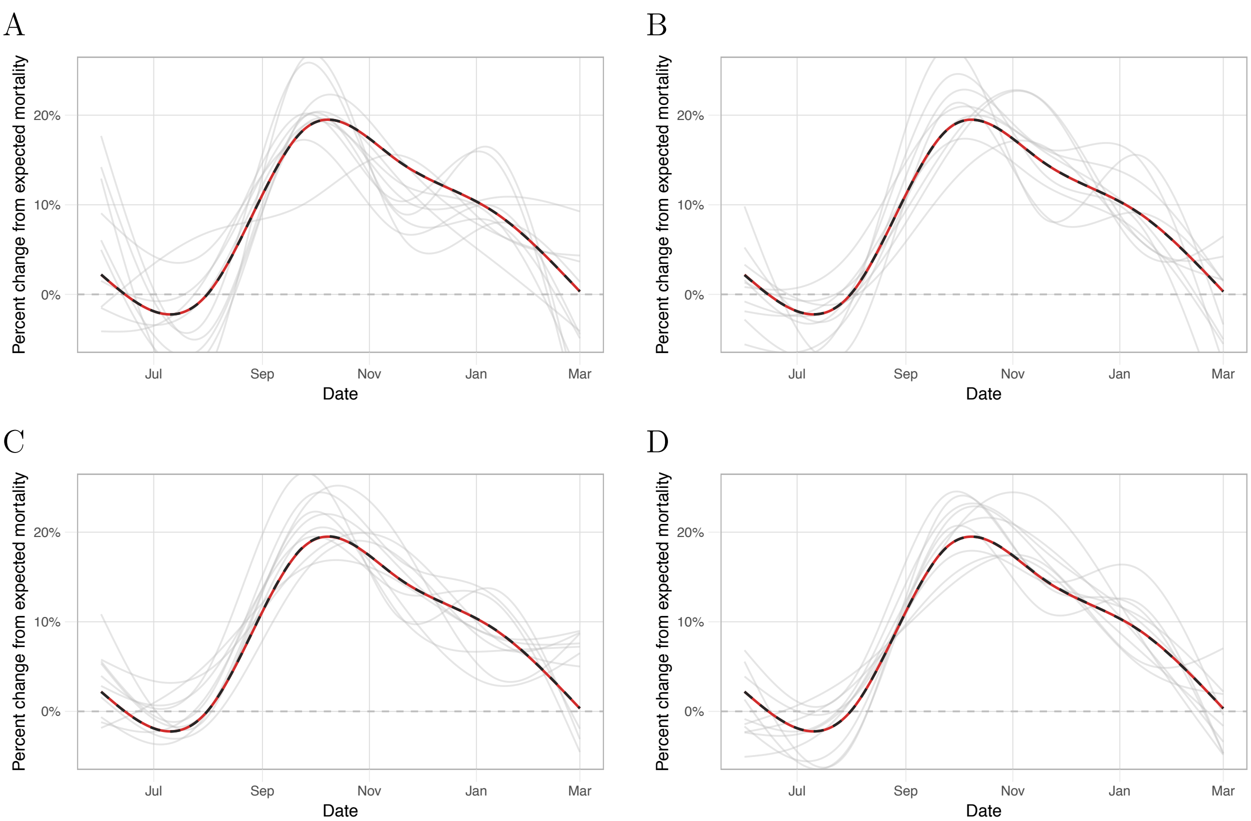
Supplementary Figure 1: Assessment of our procedure based on a simulation study. A) The solid-red curve represents the true event effect, $f(t)$, for the natural disaster scenario. The dashed-black curve is the Monte Carlo approximation of $\mathbb{E}[\hat{f}(t)]$: $1/B \sum_{b=1}^{B} \hat{f}^{(b)}(t)$. The grey curves are a random sample of ten event effects, $\hat{f}^{(b)}(t)$. B) As A) but for the infectious disease epidemic scenario. C) As A) but for the typical scenario. D) The solid-red curve represents the Monte Carlo approximation of the standard error $\sqrt{1/B \sum_{b=1}^{B} [\hat{f}^{(b)}(t) - 1/B \sum_{b=1}^{B} \hat{f}^{(b)}(t)]^2}$. The dashed-black curve represents the average of the standard errors across all simulations: $1/B \sum_{b=1}^{B} \hat{\text{SE}}[\hat{f}^{(b)}(t)]$. The grey curves represent a random sample of ten such standard errors: $\hat{\text{SE}}[\hat{f}^{(b)}(t)]$. E) As D) but for the infectious disease epidemic scenario. F) as D) but for the typical period scenario.



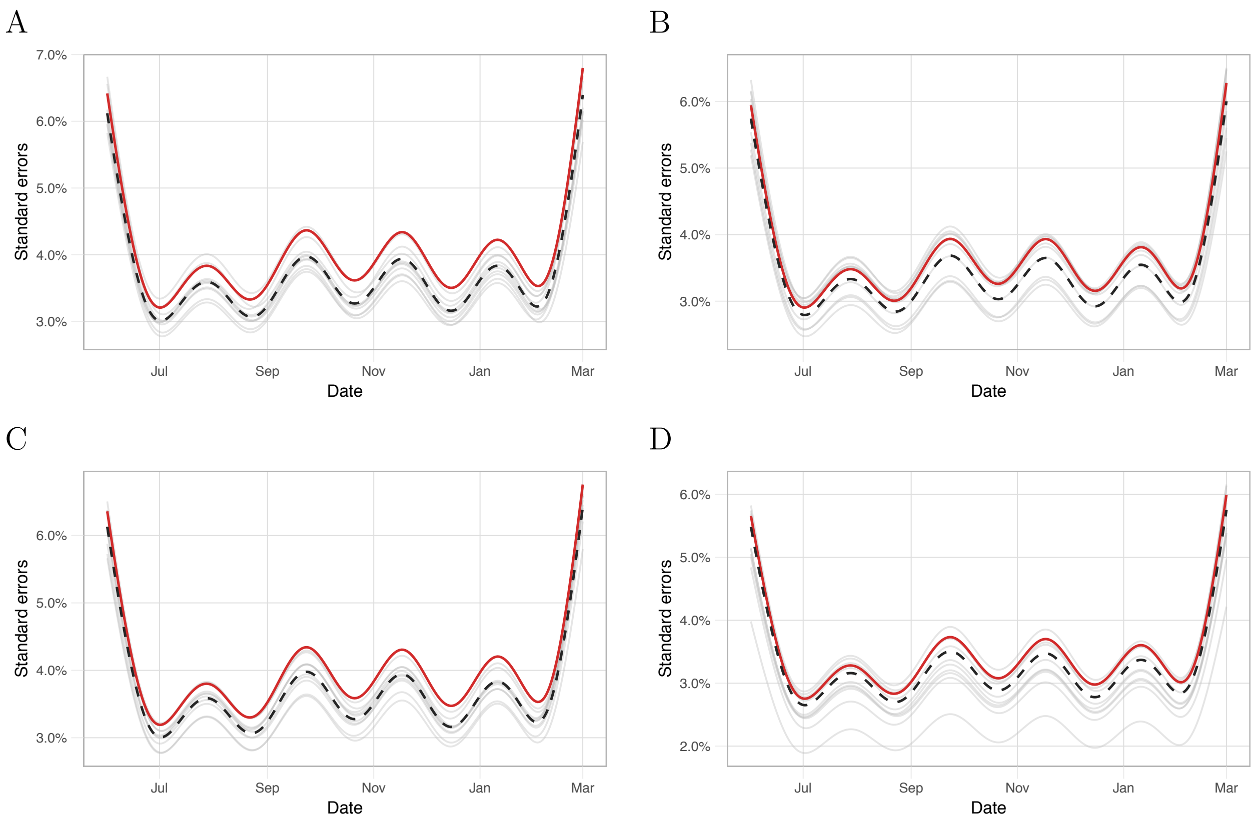
Supplementary Figure 2: Assessment of our estimate of the *event effect* based on a simulation study where the number of deaths per day was small. A) The average number of deaths per day was set to 1. The solid-red curve represents the true *event effect*, $f(t)$. The dashed-black curve is the Monte Carlo approximation of $\mathbb{E}[\hat{f}(t)]$: $1/B \sum_{b=1}^{B} \hat{f}^{(b)}(t)$. The grey curves are a random sample of ten *event effects*, $\hat{f}^{(b)}(t)$. B) As A) but for a scenario where the average number of deaths per day was set to 0.50. C) As A) but for a scenario where the average number of deaths per day was set to 0.10. D) As A) but for a scenario where the average number of deaths per day was set to 0.05.



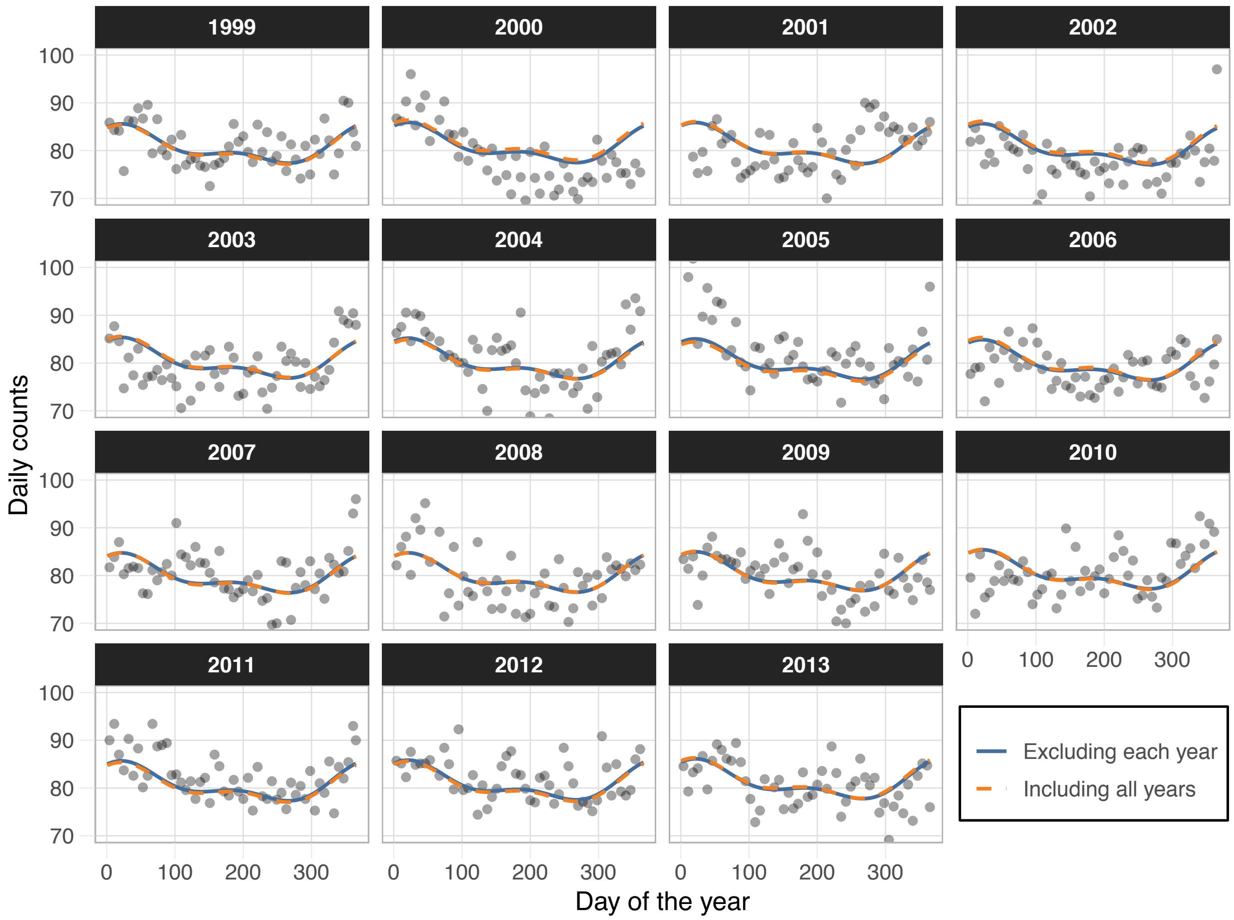
Supplementary Figure 3: Assessment of our estimate of the standard error of the *event effect* based on a simulation study where the number of deaths per day was small. A) The average number of deaths per day was set to 1. The solid-red curve represents the Monte Carlo approximation of the standard error $\sqrt{1/B\sum_{b=1}^{B}[\hat{f}^{(b)}(t) - 1/B\sum_{b=1}^{B}\hat{f}^{(b)}(t)]^2}$. The dashed-black curve represents the average of the standard errors across all simulations: $1/B\sum_{b=1}^{B}\hat{\mathrm{SE}}[\hat{f}^{(b)}(t)]$. The grey curves represent a random sample of ten such standard errors: $\hat{\mathrm{SE}}[\hat{f}^{(b)}(t)]$. B) As A) but for a scenario where the average number of deaths per day was set to 0.50. C) As A) but for a scenario where the average number of deaths per day was set to 0.10. D) As A) but for a scenario where the average number of deaths per day was set to 0.05.



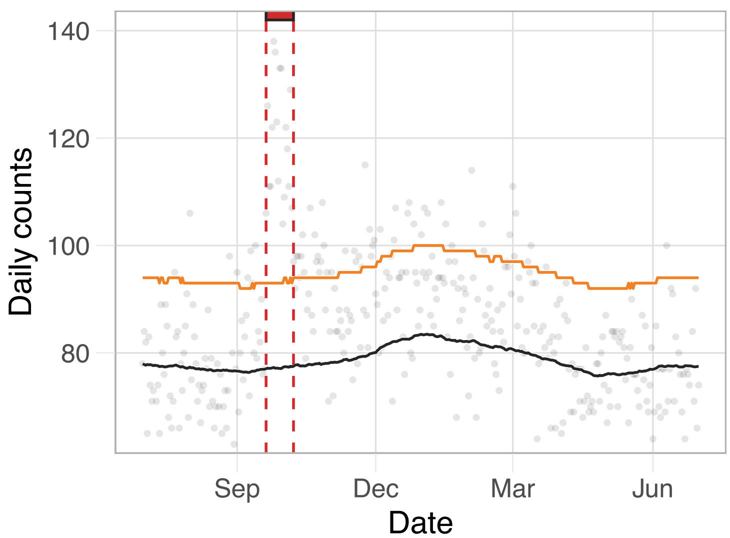
Supplementary Figure 4: Assessment of our estimate of the *event effect* based on a simulation study where we varied the length of the control region. A) The length of the control region was 8 years. The solid-red curve represents the true *event effect*, $f(t)$. The dashed-black curve is the Monte Carlo approximation of $\mathbb{E}[\hat{f}(t)]$: $1/B \sum_{b=1}^{B} \hat{f}^{(b)}(t)$. The grey curves are a random sample of ten *event effects*, $\hat{f}^{(b)}(t)$. B) As A) but for a scenario where the length of the control region was 6 years. C) As A) but for a scenario where the length of the control region was 4 years. D) As A) but for a scenario where the length of the control region was 2 years.



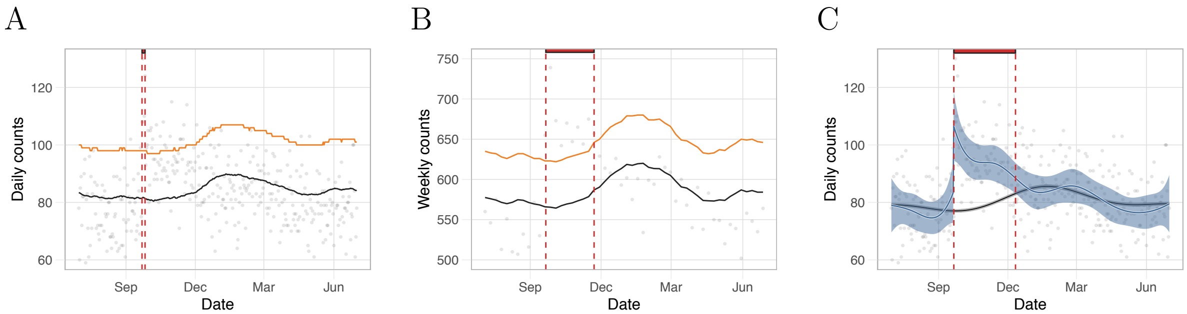
Supplementary Figure 5: Assessment of our estimate of the standard error of the *event effect* based on a simulation study where we varied the length of the control region. A) The length of the control region was 8 years. The solid-red curve represents the Monte Carlo approximation of the standard error $\sqrt{1/B\sum_{b=1}^{B}[\hat{f}^{(b)}(t) - 1/B\sum_{b=1}^{B}\hat{f}^{(b)}(t)]^2}$. The dashed-black curve represents the average of the standard errors across all simulations: $1/B\sum_{b=1}^{B}\hat{\text{SE}}[\hat{f}^{(b)}(t)]$. The grey curves represent a random sample of ten such standard errors: $\hat{\text{SE}}[\hat{f}^{(b)}(t)]$. B) As A) but for a scenario where the length of the control region was 6 years. C) As A) but for a scenario where the length of the control region was 4 years. D) As A) but for a scenario where the length of the control region was 2 years.



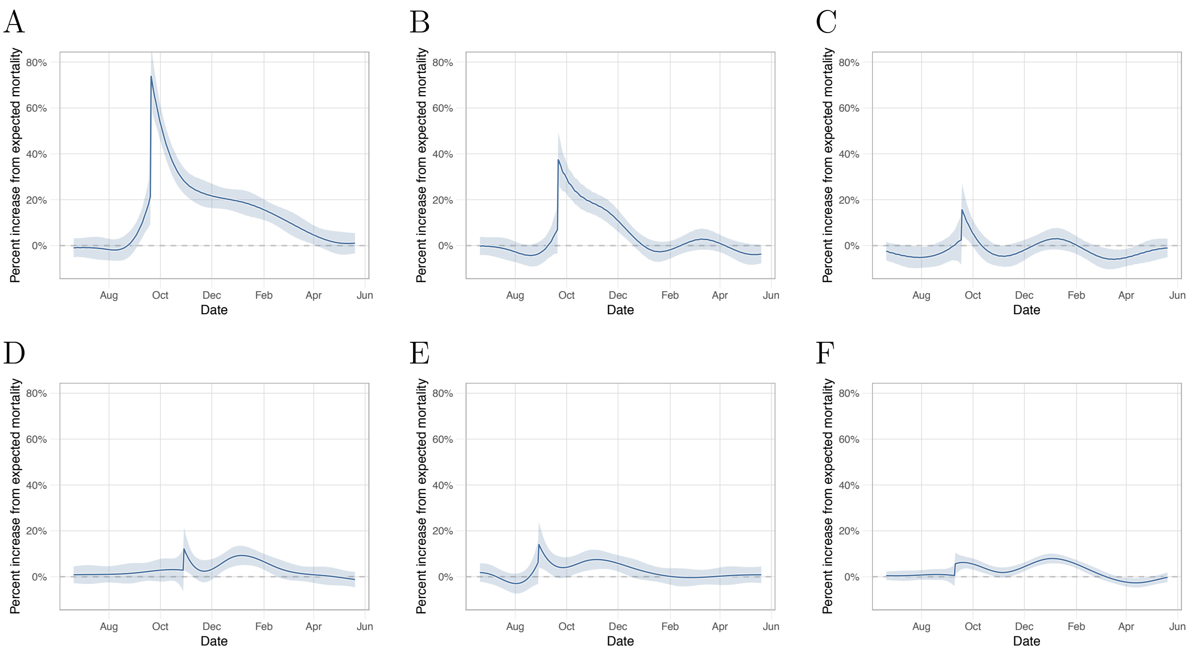
Supplementary Figure 6: Cross-validation study to assess the performance of our mean model. For years 1999 to 2013 in Puerto Rico we removed each year, one by one, estimated μ_t without that year and compared it to the estimate obtained when including that year in the analysis. The title of each pane represents the year that was removed. The points are the average deaths for every week , the solid-blue curve is the expected value when excluding each year, and the dashed-orange curve is the expected value when we did not exclude each year.



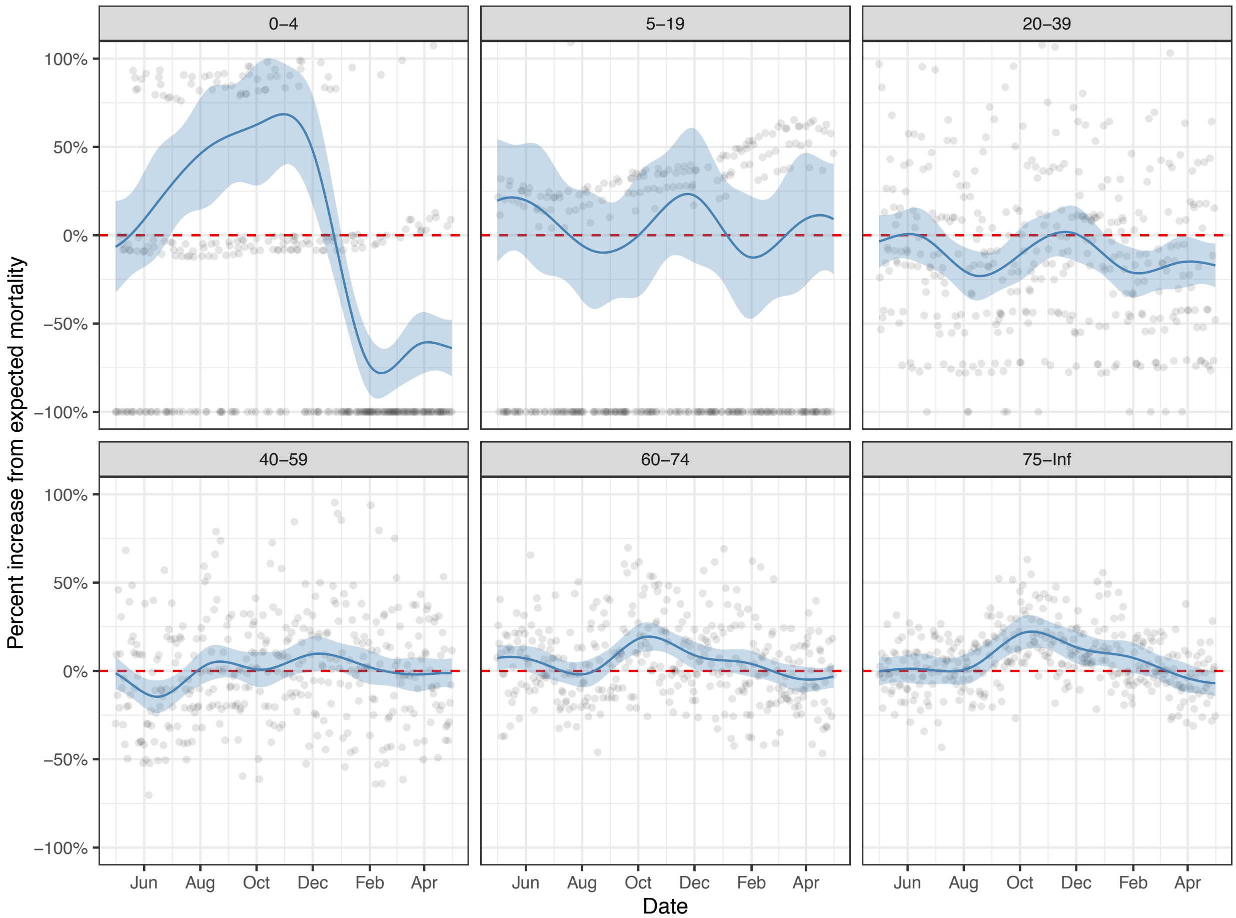
Supplementary Figure 7: Farrington model fit to daily Puerto Rico data for a period that includes the landfall of Hurricane Maria. Gray points represent daily deaths counts. The black and the orange curves are the expected number of daily counts and the threshold for significant excess deaths, respectively, as defined by the Farrington algorithm. The red rectangle denotes the number of consecutive days with excess deaths since the landfall of Hurricane Maria as determined by the Farrington algorithm.



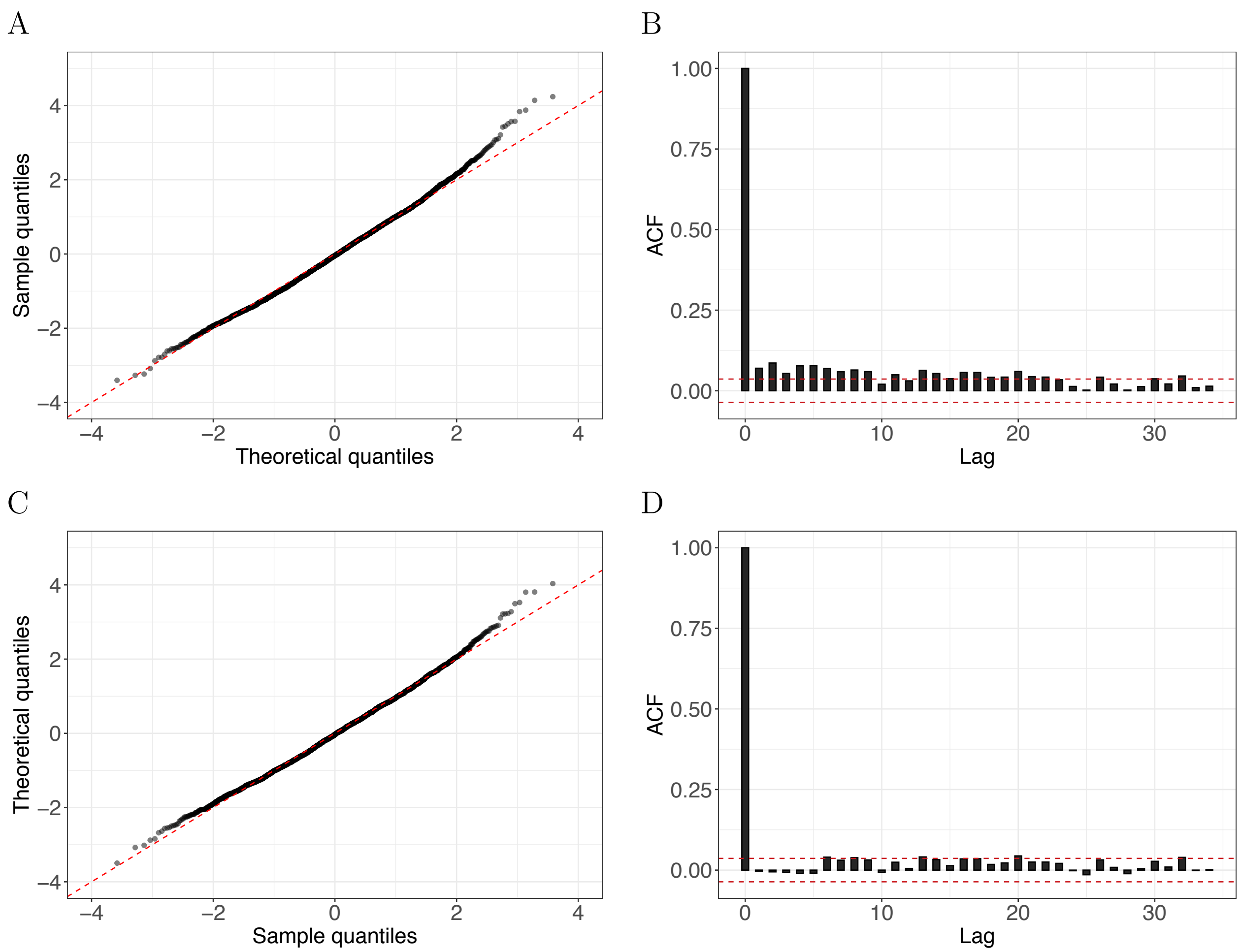
Supplementary Figure 8: Comparison between the Farrington model and our method based on estimates for Puerto Rico from a period including Hurricane Georges. A) Gray points represent daily deaths counts. The black and the orange curves are the expected number of daily counts and the threshold for significant excess deaths, respectively, as defined by the Farrington algorithm. The red rectangle denotes the number of consecutive days with excess deaths since the landfall of Hurricane Maria as determined by the Farrington algorithm. B) Gray points represent weekly death counts. The black and the orange curves are the expected number of daily counts and the threshold for significant excess deaths, respectively, as defined by the Farrington algorithm. The red rectangle denotes the number of consecutive week with excess deaths since the landfall of Hurricane Maria as determined by the Farrington algorithm. C) Gray points represent daily death counts. The black curve is the estimated expected counts based on our method and the blue curve represents the *event effect* estimate, $\hat{\mu}_t[1 + \hat{f}(t)]$. The black and blue ribbons are point-wise 95% confidence intervals for the expected counts and *event effect*, respectively. Finally, the red rectangle is as in B) but for our method.



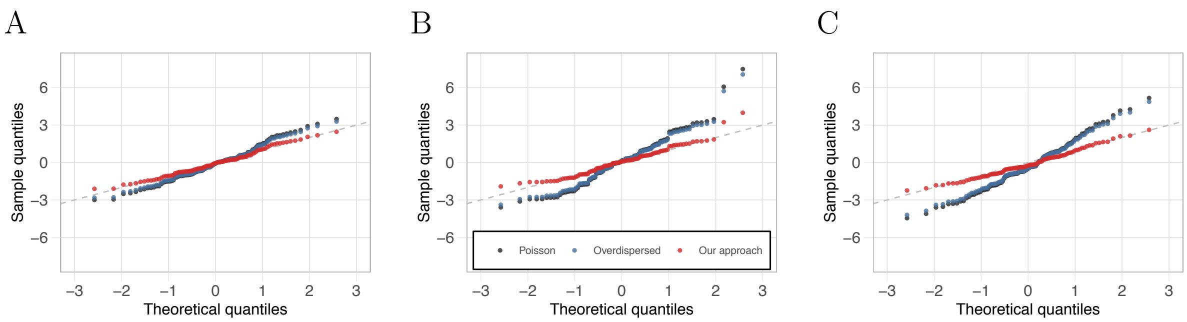
Supplementary Figure 9: Estimated hurricane effects as percent increase over expected mortality for the six hurricanes. A) *Event effect* of Hurricane Maria in Puerto Rico. The blue curve and ribbon represent the *event effect*, $100 \times f(t)$, and corresponding point-wise 95% confidence intervals. B) As A) but for Hurricane Georges in Puerto Rico. C) As A) but for Hurricane Hugo in Puerto Rico. D) As A) but for Hurricane Sandy in New Jersey. E) As A) but for Hurricane Katrina in Louisiana. F) As A) but for Hurricane Irma in Florida.



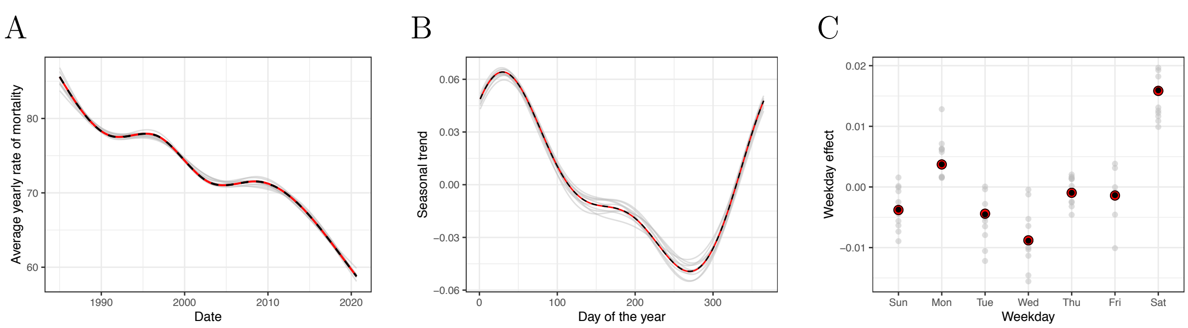
Supplementary Figure 10: Estimated effects as percent increase over expected mortality during the Chikungunya epidemic for different age groups. The grey data points correspond to observed daily percent changes from expected mortality. The blue curve and ribbon represent the event effect and its point-wise 95% confidence interval.



Supplementary Figure 11: Evidence of correlated errors. A) A Poisson GLM was fitted to Puerto Rico daily death counts of individuals 75 years and older from an interval with no known natural disasters or outbreaks (Jan 1, 2006 to Dec 31, 2013). The plot shows the Pearson residual quantiles versus theoretical quantiles from the normal distribution. One can see that the tail of the empirical data are larger than the theoretical values. B) The sample autocorrelation function for these Pearson residuals with the red-dash lines represent a 95% confidence interval centered at zero. C) As A) for residuals that were adjusted for the correlation in the data based on an estimate of the covariance matrix. D) As B) for residuals that were adjusted for the correlation in the data based on an estimate of the covariance matrix.



Supplementary Figure 12: Accounting for correlation in the error structure improves uncertainty estimates. A) Wald-statistics versus theoretical quantiles for the standard normal distribution for excess deaths estimated based on 10 days in the control interval. B) As A) but for 50 day intervals. C) As A) but for 100 day intervals.



Supplementary Figure 13: Assessment of the mean model based on a simulation study. A) The solid-red curve represents the true $\alpha(t)$, the dashed-black curve is the Monte Carlo approximation of the expected value of this estimate: $1/B \sum_{b=1}^{B} \hat{\alpha}^b(t)$, and the grey curves are a random sample of 10 $\hat{\alpha}^b(t)$s. B) As A) but for $s(t)$. C) The red points represent the true $w(t)$, the black points are the Monte Carlo approximation of the expected value of these estimates: $1/B \sum_{b=1}^{B} \hat{w}^b(t)$, and the grey points are a random sample of 10 $\hat{w}^b(t)$s.