Analysis of daily death data during the Hazelwood mine fire

Purpose

The purpose of this document is to answer the four queries below received via e-mail on 7 October 2015.

1. the parameter estimates specified in the statistical model on page 2, together with their standard error and 95% credibility intervals;
2. full details of the "natural spline with three degrees of freedom" fitted with respect to daily maximum temperature, and temperature lag, referred to on page 2;
3. With respect to the results presented in Table 1 on page 3, the results for each individual year from 2009–2013 for the period of the Mine Fire, in comparison to 2014; and
4. the results for all 6 postcodes analysed (as was done in your earlier reports).

1. The parameter estimates

Here is the complete statistical model.

$$\begin{aligned}
d_{i,t} &\sim \text{Poisson}(\mu_{i,t}), \qquad i = 1, \ldots, 4,\ t = 1, \ldots, 2191, \\
\log(\mu_{i,t}) &= \log(\text{pop}_{i,t}/10000) + \alpha_0 + \text{postcode}_i + \text{trend}_t + \text{season}_t + \text{weekday}_t \\
&+ \text{temperature}_t + \text{fire}_t, \\
\text{postcode}_i &\sim N(0, \sigma^2) \\
\text{trend}_t &= \text{ns}(\alpha_{1:2}, t, 2), \\
\text{season}_t &= \alpha_3 \cos\left[2\pi f(t)\right] + \alpha_4 \sin\left[2\pi f(t)\right], \\
\text{weekday}_t &= \alpha_{5:10} \mathbf{D}_t, \\
\text{temperature}_t &= \text{ns}(\alpha_{11:19}, \text{maximum temperature}_t, 3 \times 3), \\
\text{fire}_t &= \begin{cases} \alpha_{20}, & \text{if date}_t \in \{\text{9-Feb-2014}, \text{10-Feb-2014}, \ldots, \text{26-Mar-2014}\}, \\ 0, & \text{otherwise.} \end{cases}
\end{aligned}$$

The parameter estimates are in Table 1. The 'Label' column is the label used in the equations above (Greek alpha). I have provided the standard deviation rather than the standard error of the mean. This is because Bayesian estimates are based on a large number of Markov chain Monte Carlo samples and use an entire distribution, hence the standard deviation is a better measure of spread [1]. Standard statistical methods to calculate confidence intervals use a formula that includes the standard error of the mean.

A minor point, the correct term is 'credible interval' not 'credibility interval'.

Table 1: Model of daily deaths. Statistics are the mean, standard deviation (SD) and lower and upper 95% credible interval. Estimates are on a log scale.

	Label	Mean	SD	Lower	Upper
Intercept	α_0	−1.601	0.065	−1.732	−1.475
Trend, 1	α_1	−0.125	0.113	−0.346	0.096
Trend, 2	α_2	0.137	0.062	0.016	0.258
Season, cos	α_3	0.105	0.083	−0.057	0.269
Season, sin	α_4	0.059	0.048	−0.033	0.153
Monday	α_5	−0.069	0.064	−0.196	0.056
Tuesday	α_6	−0.096	0.065	−0.223	0.031
Wednesday	α_7	−0.042	0.063	−0.165	0.083
Thursday	α_8	−0.060	0.064	−0.186	0.064
Friday	α_9	0.049	0.063	−0.074	0.172
Saturday	α_{10}	0.008	0.063	−0.114	0.131
Temperature, 1	α_{11}	0.103	0.068	−0.030	0.238
Temperature, 2	α_{12}	−0.046	0.169	−0.378	0.286
Temperature, 3	α_{13}	−0.097	0.116	−0.324	0.133
Temperature, 4	α_{14}	−0.104	0.044	−0.193	−0.018
Temperature, 5	α_{15}	0.030	0.104	−0.176	0.228
Temperature, 6	α_{16}	0.028	0.076	−0.123	0.175
Temperature, 7	α_{17}	0.029	0.057	−0.085	0.140
Temperature, 8	α_{18}	−0.177	0.136	−0.439	0.090
Temperature, 9	α_{19}	−0.187	0.094	−0.372	−0.004
Fire	α_{20}	0.281	0.120	0.033	0.504
Postcode, 3825	postcode_1	0.285	0.031	0.225	0.346
Postcode, 3840	postcode_2	0.129	0.034	0.062	0.194
Postcode, 3842	postcode_3	−0.310	0.059	−0.426	−0.196
Postcode, 3844	postcode_4	−0.104	0.031	−0.165	−0.042

2. Details of the spline

A spline is a method of fitting a non-linear association between an exposure and outcome. In this case the exposure is daily temperature and the outcome is daily deaths. Non-linear means that the association is not a straight line, and this is needed here because both low and high temperatures are often associated with an increased risk of death. This means the association is often J- or U-shaped [2].

Another important consideration is that there can be delay between exposure to temperature and death. For example, a person exposed to low temperatures may become sick, be hospitalised and then die, and this chain of events may take a week or longer. I assumed that the delayed association was also non-linear, because previous studies have often found a strong short-term association for high temperatures, and longer lasting effect for low temperatures [3, 4, 5]. The maximum lag (delay between exposure and death) was 21 days, and this was chosen based on recent published papers and biological plausibility.

The spline was fitted using the ‘dlnm’ package in R [6]. I used three degrees of freedom as this corresponds to two change-points in the association, and this matches the theory of a change in risk for low and high temperatures. More degrees of freedom would allow a bendier association with more change-points. The change-points are partly determined by the knots which act like pivot-points. The knots were at 16.5 and 22.3 degrees C, which are the 33rd and 66th percentiles of temperature. The knot for lag was at 10.5 days, half way between 0 and 21 days (the minimum and maximum lags). The knots were selected using the default settings in 'dlnm'. The reference temperature was 20.5 degrees which is the average daily maximum temperature. The relative risk will be 1 for 20.5 degrees and all other temperatures will be compared to this average temperature.

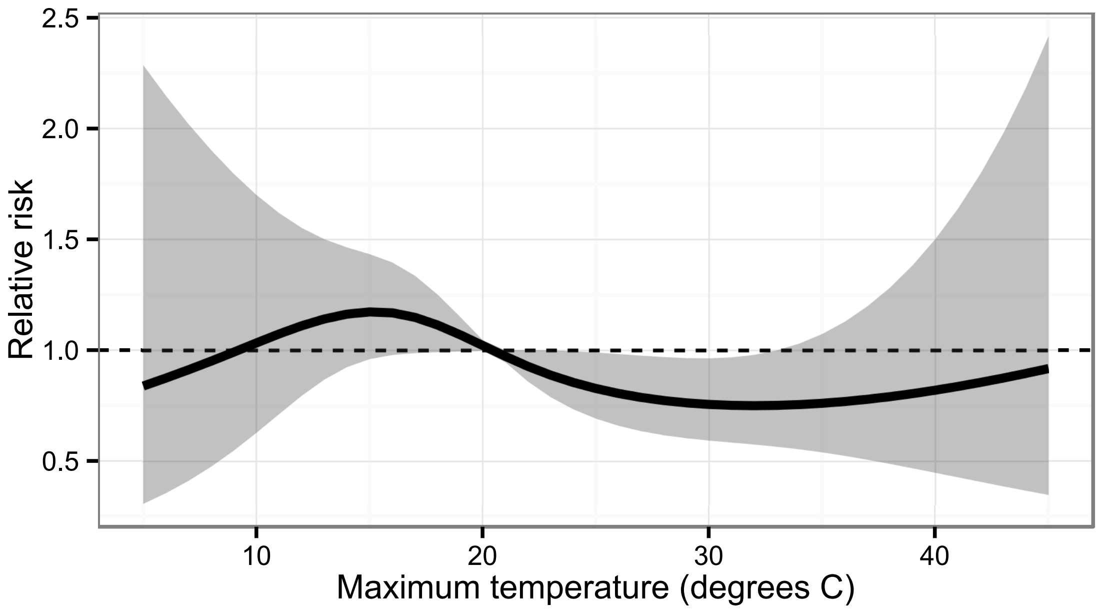

Figure 1: Estimated overall relative risk of maximum daily temperature (°C). The black line is the mean risk and the shaded areas are 95% credible intervals. The dotted horizontal line at a relative risk of 1 corresponds to no change in risk.

The overall effect of temperature is plotted in Figure 1. The lowest mean risk is around 32 degrees and the highest mean risk is around 15 degrees. The credible intervals are wider for very low and high temperatures due to the smaller number of days with extreme temperatures which increases the uncertainty.

Three estimates of the lagged effect of temperature are plotted in Figure 2. The relative risks were close to 1 for low temperatures of 10 degrees. The most notable feature is a short-term increase in risk for high temperatures (40 degrees) at lags 0 to 5 days, followed by a decrease in risk at 15 to 21 days. This decrease in risk may be due to 'harvesting' where some of the deaths caused by high temperatures were in already ill people who would have died soon after regardless of the temperature [2].

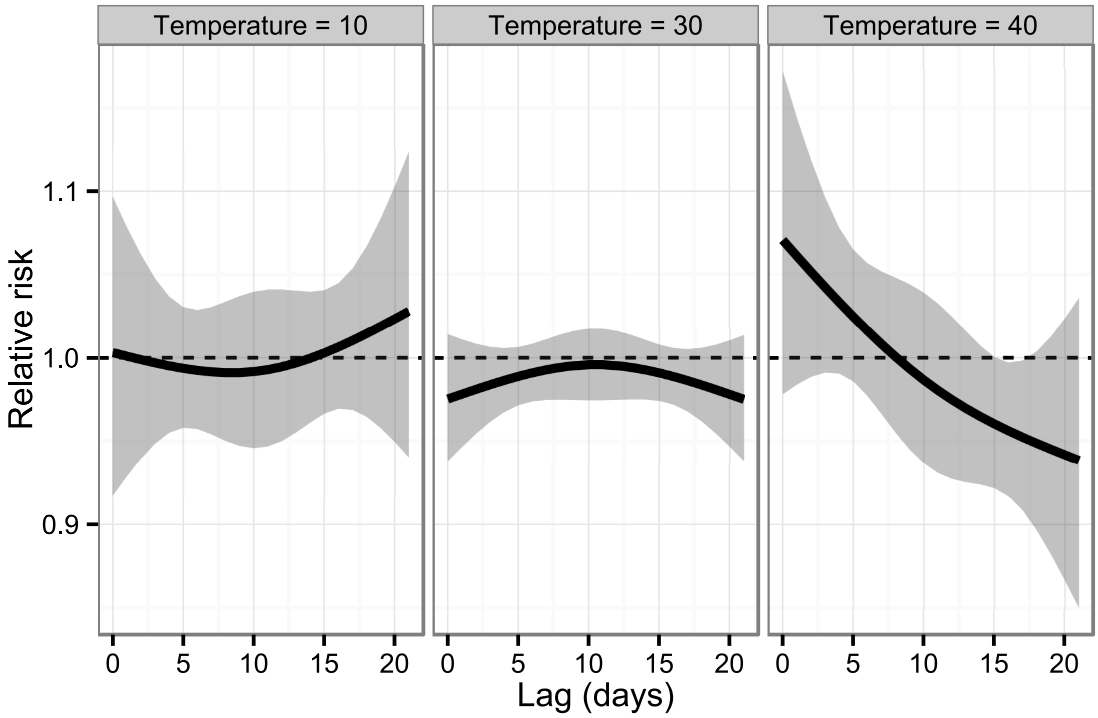

Figure 2: Estimated delayed relative risk of maximum daily temperature for three temperatures. The black line is the mean risk and the shaded areas are 95% credible intervals. The dotted horizontal line at a relative risk of 1 corresponds to no change in risk.

3. Results for individual years

The table below gives summary statistics on the daily number of deaths for the period of the fire (9 February to 26 March) in each year and postcode.

			Deaths			
Postcode	Year	N	Mean	SD	Min	Max
Churchill	2009	46	0.152	0.36	0	1
	2010	46	0.065	0.25	0	1
	2011	46	0.043	0.21	0	1
	2012	47	0.043	0.20	0	1
	2013	46	0.087	0.28	0	1
	2014	46	0.130	0.40	0	2
Moe	2009	46	0.391	0.49	0	1
	2010	46	0.500	0.75	0	2
	2011	46	0.500	0.62	0	2
	2012	47	0.511	0.66	0	2
	2013	46	0.522	0.69	0	2
	2014	46	0.717	0.81	0	3
Morwell	2009	46	0.652	0.87	0	3
	2010	46	0.261	0.49	0	2
	2011	46	0.348	0.60	0	2
	2012	47	0.426	0.50	0	1
	2013	46	0.370	0.71	0	3
	2014	46	0.413	0.62	0	2
Traralgon	2009	46	0.587	0.72	0	2
	2010	46	0.522	0.69	0	3
	2011	46	0.457	0.81	0	4
	2012	47	0.511	0.62	0	2
	2013	46	0.391	0.61	0	2
	2014	46	0.652	0.87	0	3

4. Results for all six postcodes

It is not possible to present the daily results for all six postcodes as the only daily data I have are for Moe (3825), Churchill (3842), Traralgon (3844) and Morwell (3840).

References

[1] A. Gelman, J.B. Carlin, H.S. Stern, D.B. Dunson, A. Vehtari, and D.B. Rubin. *Bayesian Data Analysis, Third Edition*. Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis, 2013.

[2] Ben Armstrong. Models for the relationship between ambient temperature and daily mortality. *Epidemiology*, 17(6):624–631, 2006.

[3] Jun Yang, Chun-Quan Ou, Yan Ding, Ying-Xue Zhou, and Ping-Yan Chen. Daily temperature and mortality: a study of distributed lag non-linear effect and effect modification in Guangzhou. *Environmental Health*, 11:63, 2012.

[4] Marco Morabito, Alfonso Crisci, Marco Moriondo, Francesco Profili, Paolo Francesconi, Giacomo Trombi, Marco Bindi, Gian Franco Gensini, and Simone Orlandini. Air temperature-related human health outcomes: Current impact and estimations of future risks in central Italy. *Science of The Total Environment*, 441:28–40, 2012.

[5] Antonio Gasparrini and et al. Mortality risk attributable to high and low ambient temperature: a multicountry observational study. *The Lancet*, 386(9991):369–375, 2015.

[6] Antonio Gasparrini. Distributed lag linear and non-linear models in R: the package dlnm. *Journal of Statistical Software*, 43(8):1–20, 2011.