From:	Justine Stansen
To:	"Bruce Armstrong"
Subject:	Hazelwood Mine Fire Inquiry
Date:	Thursday, 17 September 2015 11:28:00 AM
Attachments:	image001.jpg Death.Analysis.3.pdf

Dear Bruce

I trust you are well. We have received some further analysis undertaken by Associate Professor Adrian Barnett since the Hazelwood Inquiry hearings held earlier this month which is based on daily death data rather than monthly data. I was wondering whether you could consider the **attached** analysis and contact me to discuss your thoughts about it. The Board would be grateful for your additional input in relation to this issue.

I look forward to hearing from you.

Justine Stansen
Principal Legal Advisor
Hazelwood Mine Fire Inquiry
[redacted]

Analysis of daily death data during the Morwell mine fire

Summary

This latest analyses gives a 99% probability of an increase in deaths during the 45 days of the fire, with an estimated 23 additional deaths. This is larger than the 79% to 89% probability and 10 to 14 additional deaths from my two previous analysis. This increase in probability and deaths occurred because this analysis used daily data whereas the previous analyses used monthly data. Using days instead of months reduces the measurement error between exposure and death, and an increased statistical significance and risk is entirely expected based on the theory of measurement error [1]. This analysis also had a better control for the potential confounder of temperature, as temperature was also modelled on a daily time scale.

Introduction

This document contains my third analysis of the Morwell mine fire data. This is an updated analysis using daily death data for four postcodes for the years 2009 to 2014.

Methods

Data

The death data were daily numbers from 1 January 2009 to 31 December 2014, which is 2191 days. The deaths were split by four postcodes (3840-Morwell, 3842-Churchill, 3825-Moe, 3844-Traralgon) according to usual place of residence. There were 3,414 deaths in total.

I used population data from the Australian Bureau of Statistics for each postcode over time. This is a further improvement on my previous analyses which used overall population data for the Latrobe Valley.

The temperature data came from the Bureau of Meteorology weather station at Morwell (station number: 85280), which provided daily maximum temperature. Two days were missing and I imputed the missing temperature using the mean temperature for the days either side of the missing day. I used maximum temperature rather than mean or minimum temperature because previous research found that most common temperature measures are highly correlated and perform equally well when predicting daily death rates [2].

Statistical methods

I used a regression model to examine the key hypothesis of whether deaths rates were higher during the 45 days of the fire.

I give the model as an equation below and then explain each line of the equation.

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

The index i is for postcode and the index t is for time. I used a Poisson model as the data are daily counts of deaths. The trend was fitted as a natural spline (ns) with two degrees of freedom which allowed the underlying death rate to change slowly during 2009 to 2014 due to factors such as an ageing population. Season was fitted as an annual sinusoid and f is the fraction of the year from 0 (1 January) to 1 (31 December) [3]. I modelled the expected small difference in death rates by day of the week using an independent effect on each day with Sunday as a the reference day.

Temperature was modelled as a non-linear variable to allow for increased risks in low and high temperatures [4]. To allow for the known delay between exposure to temperature and death I also included a lag with a delay up to 21 days. Both temperature and lag were fitted using a natural spline with three degrees of freedom which is large enough to model a non-linear association.

To check the adequacy of the model I examined the residuals (difference between observed and predicted) using a histogram and autocorrelation plot.

Results

Simple table

Table 1 shows a higher mean number of daily deaths in all four postcodes during the period of the fire compared with all other times. These crude figures do not adjust for the seasonal pattern in deaths, and the regression model below should give a truer picture of any increase in death rates.

Table 1: Summary statistics on daily deaths by postcode and the time of the fire using data for 1 January 2009 to 31 December 2014

			Deaths			
Postcode	Fire	N	Mean	SD	Min	Max
Churchill	No	2145	0.075	0.27	0	2
	Yes	46	0.130	0.40	0	2
Moe	No	2145	0.558	0.74	0	5
	Yes	46	0.717	0.81	0	3
Morwell	No	2145	0.396	0.63	0	4
	Yes	46	0.413	0.62	0	2
Traralgon	No	2145	0.522	0.73	0	6
	Yes	46	0.652	0.87	0	3
All	No	8580	0.388	0.65	0	6
	Yes	184	0.478	0.73	0	3

Plots of daily deaths over time

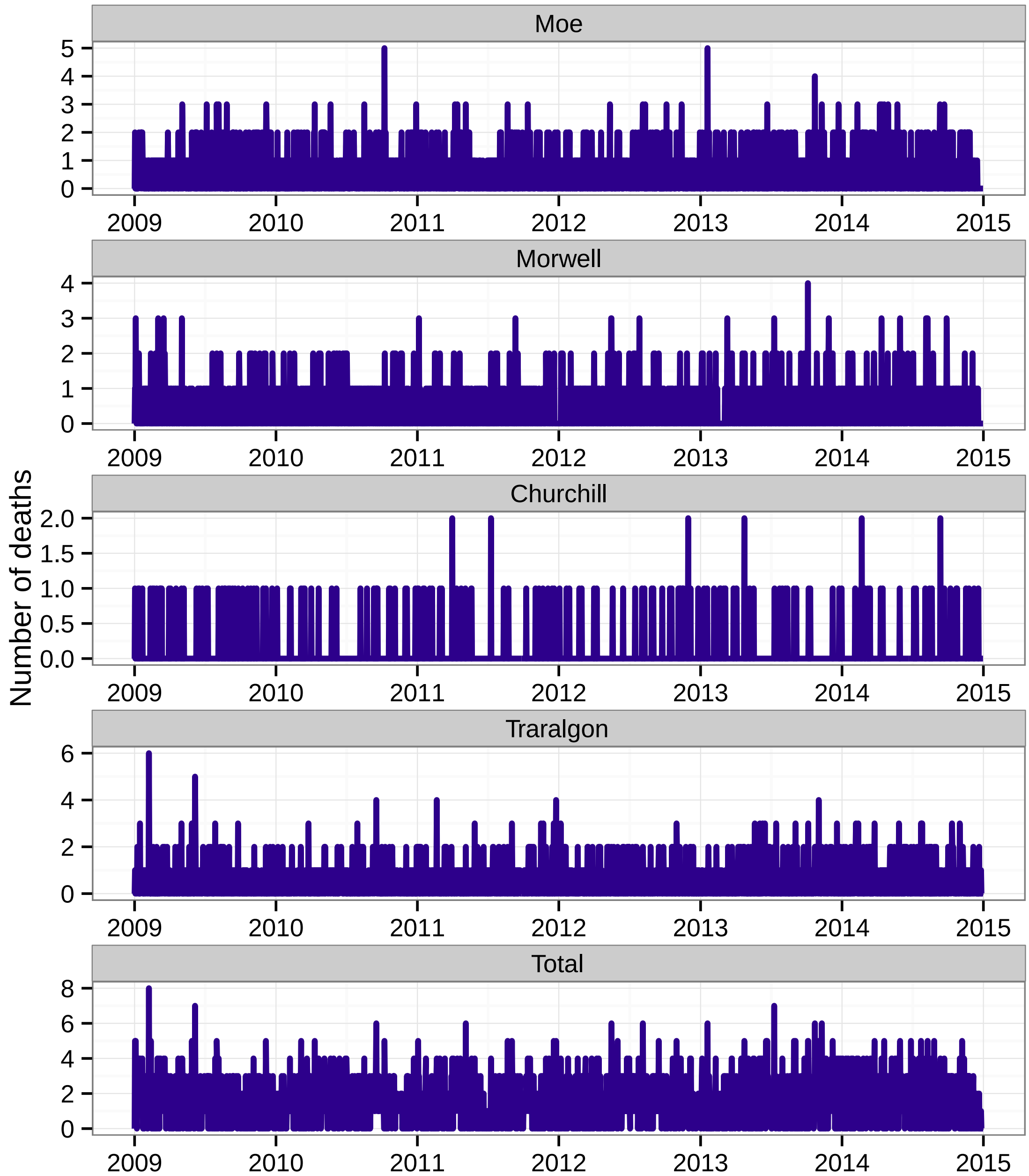

Figure 1: Daily death numbers in each postcode and the total number of deaths across the four postcodes for 1 January 2009 to 31 December 2014.

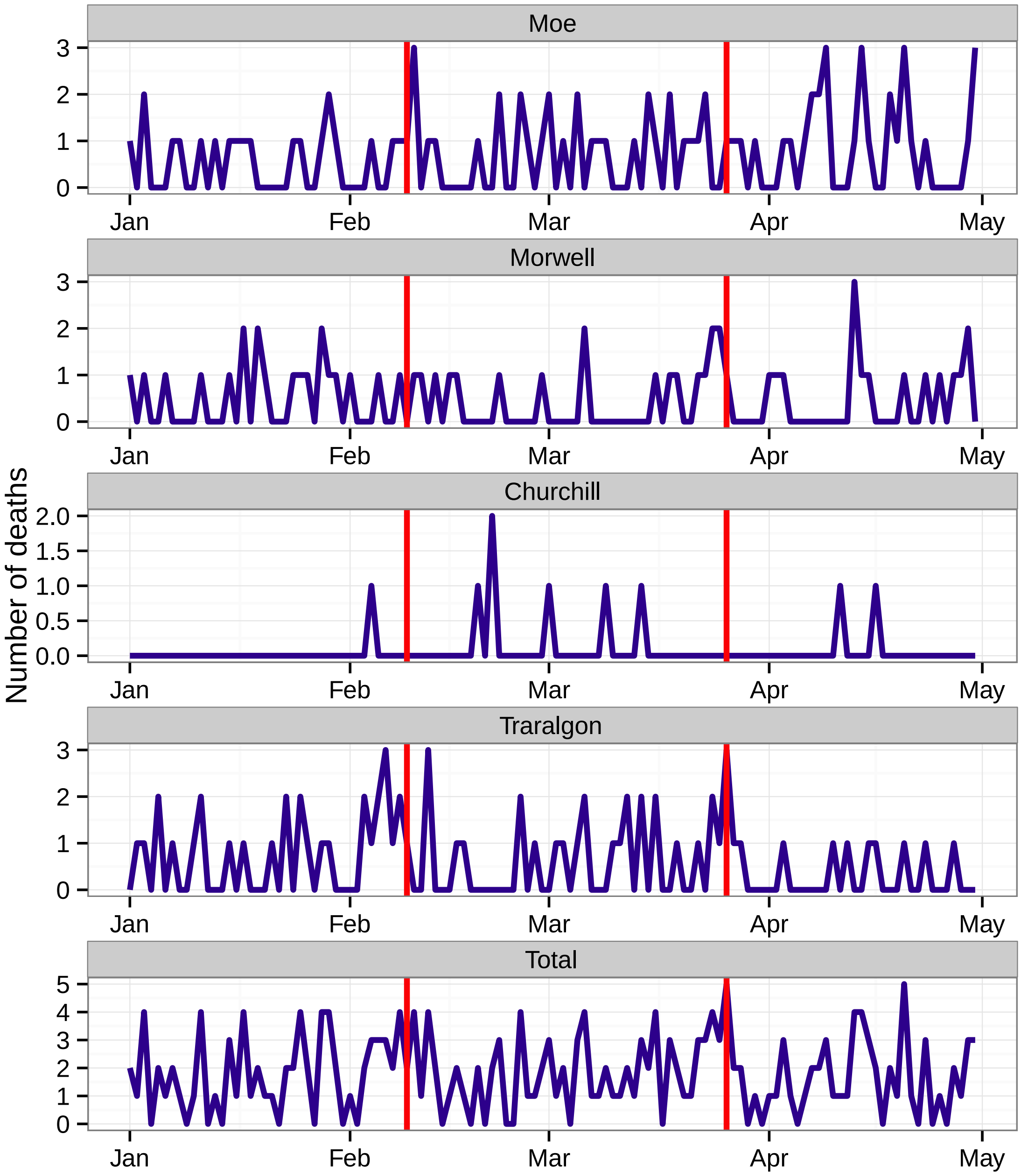

Figure 2: Daily death numbers in each postcode and the total number of deaths across the four postcodes for 1 January 2014 to 30 April 2014. The start and end of the fire are shown by vertical red lines.

Statistical model results

Table 2: Model of daily deaths. Statistics are the mean and lower and upper 95% credible interval. Estimates are on a log scale except for the relative risk and absolute number of deaths.

	Mean	Lower	Upper
Intercept	−1.601	−1.732	−1.475
Trend, 1	−0.125	−0.346	0.096
Trend, 2	0.137	0.016	0.258
Postcode, 3825	0.285	0.225	0.346
Postcode, 3840	0.129	0.062	0.194
Postcode, 3842	−0.310	−0.426	−0.196
Postcode, 3844	−0.104	−0.165	−0.042
Season, cos	0.105	−0.057	0.269
Season, sin	0.059	−0.033	0.153
Monday	−0.069	−0.196	0.056
Tuesday	−0.096	−0.223	0.031
Wednesday	−0.042	−0.165	0.083
Thursday	−0.060	−0.186	0.064
Friday	0.049	−0.074	0.172
Saturday	0.008	−0.114	0.131
Fire, relative risk	1.324	1.034	1.656
Additional deaths during fire, 3825	8.271	0.860	16.731
Additional deaths during fire, 3840	5.848	0.608	11.830
Additional deaths during fire, 3842	1.124	0.117	2.273
Additional deaths during fire, 3844	7.733	0.804	15.642
Additional deaths, all postcodes	22.976	2.388	46.476

The probability that the death rate was higher than the average during the fire is 0.99. This means that the probability that the death rate was not higher than the average during the fire is 0.01. The mean increase in deaths is 1.32 as a relative risk, or 32 as a percentage. The 95% credible interval for the relative risk does not include 1, indicating that the risk was higher than average during the fire. The mean estimated number of extra deaths during the fire over the four postcodes is 23.

Effect of temperature

The effect of temperature in Figure 3 is exactly as expected. It shows a steep rise in risk for high temperatures on the day of exposure, and smaller but longer lasting risk for low temperatures [4].

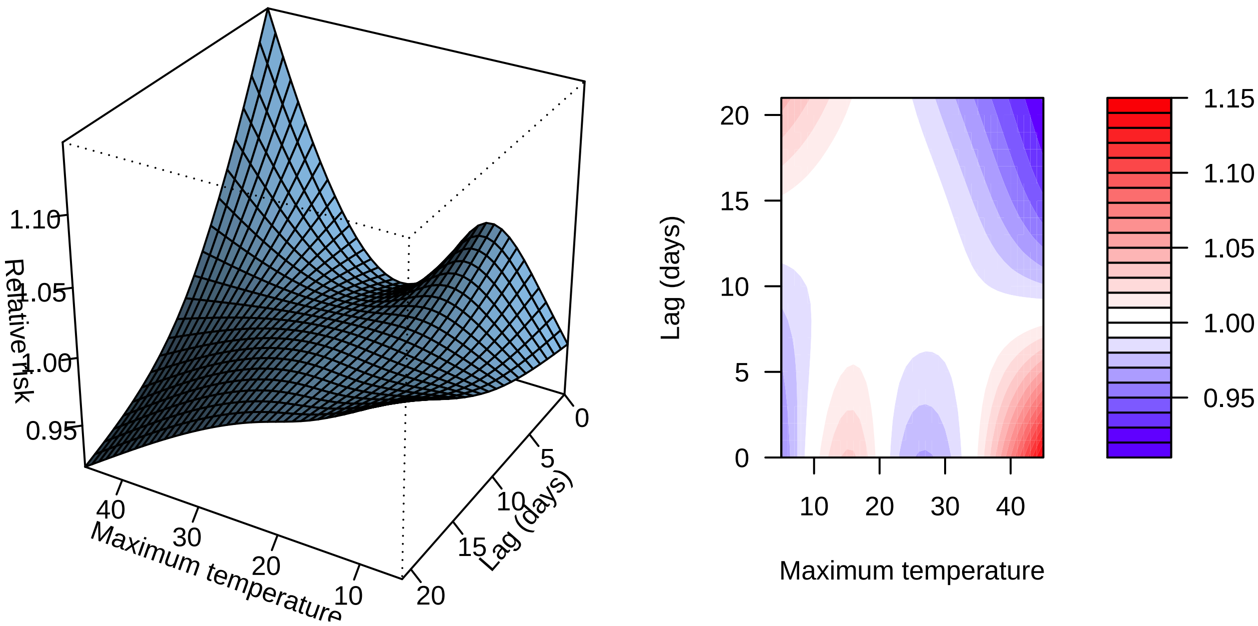

Figure 3: Estimated relative risk of maximum temperature (°C) by temperature and lag using a surface plot (left) and contour plot (right).

Residual plots

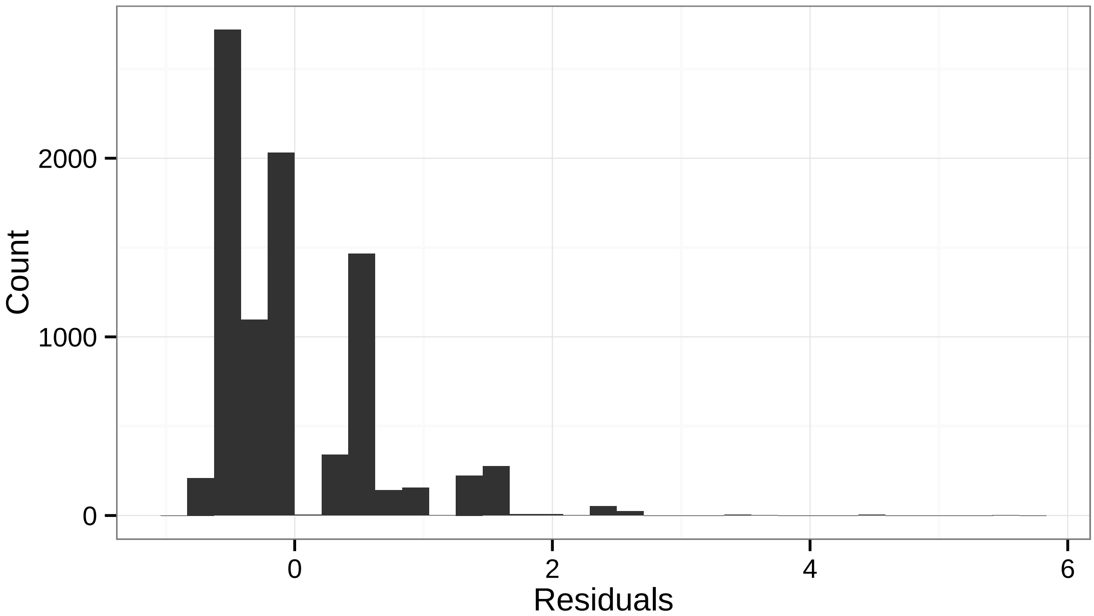

Figure 4: Residual histogram from the model of daily deaths.

The histogram of residuals are centred on zero but with a positive skew which is as expected when modelling small counts (Figure 4). There were four relatively large residuals over 4 as shown in Table 3. The large residual in Traralgon on 7/Feb/2009 may be the Black Saturday bushfires.

Table 3: Four large residuals where the model greatly under-predicted the number of deaths.

Date	Postcode	Deaths	Predicted	Residual	Pearson residual
08/Oct/2010	Moe	5	0.60	4.40	5.66
19/Jan/2013	Moe	5	0.51	4.49	6.27
07/Feb/2009	Traralgon	6	0.57	5.43	7.22
06/Jun/2009	Traralgon	5	0.58	4.42	5.78

The Pearson goodness of fit statistic is 8749 which is smaller than test limit of 8958, which is the 95th percentile of a chi-squared distribution [5]. This indicates that the model is an adequate fit to the data.

The autocorrelation plots of the residuals show no residual autocorrelation in any postcode as the correlations are small and close to zero (Figure 5). This means there is unlikely to be any residual confounding by other short-term environmental factors (e.g., humidity).

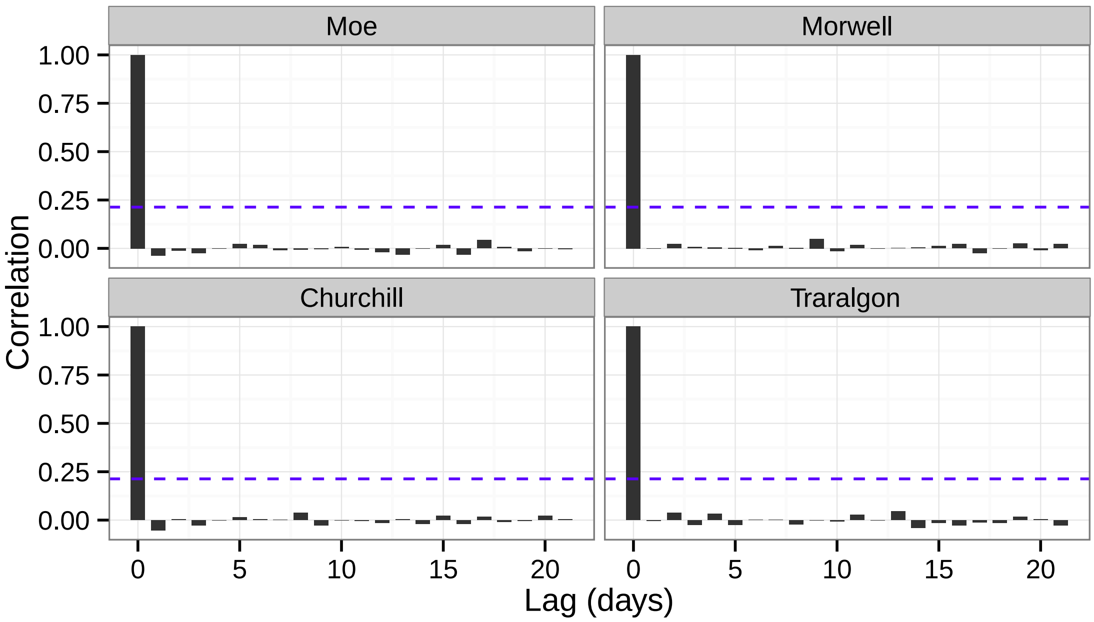

Figure 5: Autocorrelation of residuals from the model of daily deaths by postcode. The dotted horizontal blue line is the limit for assessing significant autocorrelation.

References

[1] Jennifer A Hutcheon, Arnaud Chiolero, and James A Hanley. Random measurement error and regression dilution bias. *BMJ*, 340, 2010.

[2] A. Barnett, S Tong, and ACA Clements. What measure of temperature is the best predictor of mortality? *Environmental Research*, 110(6):604–611, 2010.

[3] Adrian G Barnett and Annette J Dobson. *Analysing Seasonal Health Data*. Springer, Berlin, Heidelberg, 2010.

[4] 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.

[5] Annette J Dobson and A.G̃. Barnett. *An Introduction to Generalized Linear Models*. Texts in Statistical Science. Chapman & Hall/CRC, Boca Raton, FL, 3rd edition, 2008.

[6] Martyn Plummer. *rjags: Bayesian graphical models using MCMC*, 2013. R package version 3-11.

Appendix

JAGS Code

This is the code using the JAGS software that runs the Bayesian regression model of daily deaths [6].

```
model{
# likelihood
for (i in 1:N){
deaths[i] ~ dpois(mu[i]);
log(mu[i]) <- log.pop[i] + alpha + weekday[i] + trend[i] + gamma*fire[i]
+ delta.c[pcode[i]] + season[i] + temp[i];
weekday[i] <- inprod(dow[i,1:6], phi[1:6]);
trend[i] <- inprod(time[i,1:n.time], beta[1:n.time]);
season[i] <- theta[1]*cosw[i] + theta[2]*sinw[i];
temp[i] <- inprod(temperature[i,1:n.temp], zeta[1:n.temp]);
}
# priors
alpha ~ dnorm(0, 0.001) # intercept
for (k in 1:n.time){
beta[k] ~ dnorm(0, 0.001) # time trend
}
gamma ~ dnorm(0, 0.001) # fire
for (k in 1:6){
phi[k] ~ dnorm(0, 0.001) # week day
}
for (k in 1:n.temp){
zeta[k] ~ dnorm(0, 0.001) # temperature
}
for (k in 1:n.pcode){
delta[k] ~ dnorm(0, tau.delta); # random intercept for postcode
delta.c[k] <- delta[k] - mu.delta;
# absolute numbers
```

```
    absolute[k] <- mu.deaths[k]*(rr-1)
}
absolute[5] <- sum(absolute[1:4]) # total deaths
tau.delta ~ dgamma(1,1)
for (k in 1:2){
theta[k] ~ dnorm(0, 0.001); # season
}
## scalars
mu.delta <- mean(delta[1:n.pcode])
p.gamma <- step(gamma) # p-value for positive risk
rr <- exp(gamma) # relative risk
}
```