Analysis of daily death data during the Hazelwood mine fire

Purpose

The purpose of this document is to answer two queries from Professor Bruce Armstrong:

1. The mean number of deaths per day for each postcode. The period over which this average has been calculated is not stated; It should be. As I see it, the period should (a) be relatively recent so that it can provide a reasonably unbiased estimate of the expected number of deaths in the four postcode areas over the period of the fire, (b) not include the observed deaths during the period of the mine fire and (c) be based on a period long enough to remove most of the effect of day to day variation in daily numbers on the calculated mean numbers. All these may be true, but it is not clear that they are.
2. Exp(α_{20}), the relative risk of death during the fire. As far as I can tell this is the relative risk across all four postcodes. If this is true, postcode specific relative risks have not been used when estimating the excess deaths and, therefore, previously apparent variation between postcodes in relative risk of death during the period of the mine fire is not taken into account when calculating the numbers of excess deaths. If this is correct, a deficit of deaths in Morwell during the period of the mine fire would be obscured in this analysis.

Summary response

1. I tried a few alternative methods for calculating the mean number of deaths based on Professor Armstrong's suggestions. The estimated number of deaths during the fire were similar regardless of which mean was used.
2. A model using postcode specific relative risks was not as good a fit to the data as a model with a common relative risk. Hence the previous results using a common relative risk should be preferred. However, even for a model with a varying risk across postcodes, there is an increased relative risk of death during the fire in Morwell.

More detailed analyses that address the two queries are given in the sections below.

1. The mean number of deaths per day for each postcode

The estimated additional number of deaths during the fire in each postcode were calculated using:

$$45 \times \overline{d}_i \times [\exp(\alpha_{20}) - 1],$$

where $\overline{d}_i$ is the mean number of daily deaths in postcode i and $\exp(\alpha_{20})$ is the relative risk of death during the fire. The daily estimate is multiplied by 45 days to give an estimate for the period of the fire.

Prof Armstrong queried the time period used to calculate the mean number of deaths ($\overline{d}_i$). This was based on the entire period of available data, from 1 January 2009 to 31 December 2014 and hence includes the period of the fire. My reasoning for using the entire period was that the influence of the fire would be relatively small given the large sample size.

However, I agree with Prof Armstrong's reasoning that the baseline mean should exclude the period of the fire, I therefore show some alternative calculations below.

Table 1: Mean number of additional deaths during the fire and 95% credible intervals using alternative versions of the baseline mean number of deaths in each postcode ($\overline{d}_i$).

Postcode	Period used to calculate the baseline mean	Baseline mean	Mean	Lower	Upper
3825	All data	0.56	8.2	0.9	16.5
3825	Period of fire in previous years (2009–2013)	0.48	7.1	0.7	14.3
3825	Period of fire in previous two years (2012–2013)	0.52	7.5	0.8	15.2
3840	All data	0.40	5.8	0.6	11.7
3840	Period of fire in previous years (2009–2013)	0.41	6.0	0.6	12.1
3840	Period of fire in previous two years (2012–2013)	0.40	5.8	0.6	11.7
3842	All data	0.08	1.1	0.1	2.2
3842	Period of fire in previous years (2009–2013)	0.08	1.1	0.1	2.3
3842	Period of fire in previous two years (2012–2013)	0.06	0.9	0.1	1.9
3844	All data	0.52	7.6	0.8	15.5
3844	Period of fire in previous years (2009–2013)	0.49	7.2	0.7	14.6
3844	Period of fire in previous two years (2012–2013)	0.45	6.6	0.7	13.3
Total	All data	1.56	22.7	2.4	46.0
Total	Period of fire in previous years (2009–2013)	1.47	21.4	2.2	43.3
Total	Period of fire in previous two years (2012–2013)	1.43	20.9	2.2	42.2

The results in Table 1 show that the alternative calculations for the baseline mean have only a minor impact on the estimated additional number of deaths. The 'period of the fire' is 9 February to 26 March.

2. Postcode specific relative risks

Prof Armstrong is correct in stating that $\exp(\alpha_{20})$ is the relative risk common to all four postcodes. My reasoning for using a common relative risk is that the previous analysis found little evidence for a postcode-specific effect (Table 4 in December 2014 analysis [1]). However, we can revisit this issue given that we are now examining daily data.

Given the time constraints of providing these analyses I could not use a Bayesian approach as these take time to run. Instead I used a standard statistical approach, and I show the similarity of the Bayesian and standard models below. The major differences between the two approaches are: i) how they estimate the model parameters, and ii) the interpretation of the parameters. Both approaches used the same model structure (e.g., same variables to control for daily temperature).

The estimates in Figure 1 are very similar for both the means and 95% intervals. The only noticeable difference is for the intercept, where the Bayesian credible interval is narrower

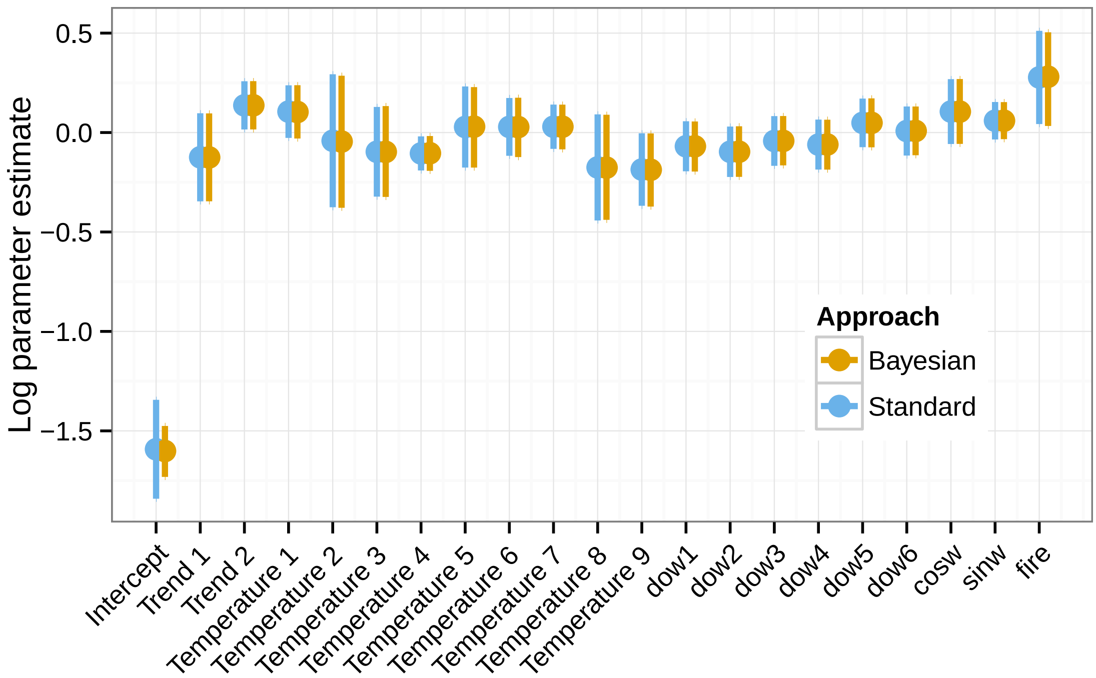

Figure 1: Comparison of parameter estimates using a standard statistical and Bayesian approach. The dots show the mean and the vertical lines are the 95% confidence/credible intervals.

than the standard confidence interval.

Table 2: Akaike information criterion (AIC) and degrees of freedom (df) comparing the two models using a standard statistical approach. The lower the AIC the better the model.

Relative risk of fire	df	AIC
Common across postcodes	22	13301
Varying across postcodes	24	13305

To compare the model fit we can use the Akaike Information Criterion (AIC) [2] as shown in Table 2. The fit was somewhat worse for the model with the varying relative risk, therefore the model with a common risk should be preferred. The degrees of freedom is essentially the number of model parameters, so the model with a varying relative risk had two extra parameters. The varying model was more complex, but did not give a better fit to the data.

The relative risks assuming a varying model are shown in Table 3. The lowest risk was in 3825 (Moe) and the highest in 3842 (Churchill), but the range in relative risks was relatively narrow and all mean risks were increased (i.e., greater than 1).

Table 3: Estimates of the mean relative risk assuming a common and varying effect of the fire across the four postcodes.

Model	Postcode	Mean relative risk
Common effect of fire		1.32
Varying effect of fire	3825	1.29
Varying effect of fire	3840	1.31
Varying effect of fire	3842	1.38
Varying effect of fire	3844	1.35

References

[1] Adrian Barnett. An updated analysis of death data during the morwell mine fire. Technical report, Queensland University of Technology, `http://eprints.qut.edu.au/81685/`, 2 2015.

[2] K.P. Burnham and D.R. Anderson. *Model Selection and Inference: A Practical Information-Theoretic Approach*. Springer New York, 2013.