Mobile safety cameras: Estimating casualty reductions and the demand for secondary healthcare

Lee Fawcett and Neil Thorpe

November 11, 2010

1 Problems with previous work

Some issues identified by Maher:

- Should really have used a "control" set of sites to develop the regression model (or Predictive Accident Model, "PAM") before applying this to the treated sites to predict the number of casualties in the 'After' period, had no camera been installed. These sites should, in some way, be representative of the treated sites (e.g. region, urban/rural, type of roads, speed limits etc), but should not *include* any of the treated sites.
- Question over how "valid" a zero–inflated model formulation actually is.
- Should combine 'serious' and 'fatal' casualties.

Other (statistical) issues:

- Trend treated in an over–simplistic way? Could model this more accurately if data from the control sites was available from other time periods.
- Dependence between predictor variables... how can we account for this?
- Using the model to forecast savings to the NHS secondary healthcare provider: think about this more carefully and use a different posterior summary of "cost".
- Is it possible to construct informative priors for the model parameters (not been done before).

2 Reminder

We have 67 treated sites. For each site j, $j = 1, \ldots, 67$, we have:

$$\begin{aligned}
y_{j,\text{before}}^{\text{slight/serious/fatal}} &= \text{number of casualties before safety cameras} \\
y_{j,\text{after}}^{\text{slight/serious/fatal}} &= \text{number of casualties after safety cameras} \\
x_{1,j} &= \text{Speed limit} \\
x_{2,j} &= \text{Average observed speed} \\
x_{3,j} &= \text{85th \% speed} \\
x_{4,j} &= \text{\% drivers over the speed limit} \\
x_{5,j} &= \text{\% drivers at least 15mph over the speed limit} \\
x_{6,j} &= \text{Traffic flow} \\
x_{7,j} &= \text{Road classification} \\
x_{8,j} &= \text{Road type}
\end{aligned}$$

Problem with comparing simple before/after figures on y: does not take into account the fact the before figures are unusually high, and even without the installation of cameras there would probably be a natural "settling down" reduction anyway (regression–to–mean, RTM). Thus, the before figure is too high, and so the effect of the cameras would be exaggerated.

Usual approach: Empirical Bayes.

- Use a statistical model to "predict" casualty frequencies at each site in the after period, had no cameras been used – then compare *this* with the after figure.
- We shouldn't ignore the before figure – this *did* happen! However, for each site j it should be 'toned down' a bit by what we'd expect to happen at other (control) sites with similar characteristics to site j.
- Use a Poisson–Gamma model:

$$\begin{aligned}
y_j &\sim \text{Poisson}(m_j) \\
m_j &\sim \text{Gamma}(\gamma, \gamma/\mu_j),
\end{aligned}$$

  i.e. a Poisson distribution for the number of casualties at each site j, but where the mean of this Poisson is itself allowed to vary according to a Gamma distribution – the mean and variance of a Gamma(a, b) is a/b and a/b^2, respectively, giving a mean and variance here of μ_j and μ_j^2/γ.
- We are thus using a Gamma "prior" distribution for the mean number of casualties with common "shape" parameter γ and "rate" parameter γ/μ_j.
- Why use a Gamma distribution? Mathematical convenience. The Gamma is the "conjugate prior" for the Poisson, and combining the two in this way gives a Gamma "posterior" distribution for m_j (i.e. the distribution for the mean number of accidents after seeing some data):

$$m_j | y_j \sim \text{Gamma}(\gamma + y_j, \gamma/\mu_j + 1).$$

- The mean of this posterior, since it is gamma, is just

$$\frac{\gamma + y_j}{\mu_j/\gamma + 1} = \alpha_j \mu_j + (1 - \alpha_j) y_j, \quad (1)$$

where $\alpha_j = \gamma/(\gamma + \mu_j)$. Nice! This is a weighted average of what actually happened in the before period (y_j) and what we'd *expect* to see at such a site (μ_j).

- How do we get γ and μ_j? Answer: regression techniques using the control data.

3 Issues with EB

- A bit artificial? Because of 'recent' advances in computer simulation techniques, no need to stick with the gamma prior for mathematical convenience. Maybe better priors to use?
- The EB approach gives the "posterior mean"; possibly not the best summary to use, especially if the posterior is highly skewed.
- Estimates for γ and μ_j, taken from the control sites, are used as the 'true values' in EB – the variability of these estimates is not incorporated into the analysis, and so EB estimates of casualty frequency will be over–optimistic in their precision (i.e. will have standard errors that are too small).
- Implementing a "Fully Bayesian" analysis can get round all of these issues, and has the potential to incorporate expert opinion into the prior distributions!

4 New analysis

4.1 Empirical Bayes

4.1.1 Step 1: Develop the PAM

Initial investigations reveal substantial dependencies between x_1 (speed limit) and some other predictor variables, and x_3 (85th %–ile speed) and other predictor variables, so we exclude these.

Using each *control* site j, we then proceed to fit a model of the form:

$$y_j = \exp\left\{\beta_0 + \beta_2 x_{j,2} + \beta_4 x_{j,4} + \ldots + \beta_8 x_{j,8}\right\},$$

for each of the slight, serious and fatal casualties separately.

- Doing so gives no appropriate model for serious nor fatal, even after transformations, so we combine all casualty severities into one class.
- Using backwards elimination, we see that variables x_5 (% at least 15mph over the speed limit) and x_8 (road type) are insignificant predictors, leaving just

$$\begin{aligned} x_{2,j} &= \text{Average observed speed at site } j \\ x_{4,j} &= \%\text{ drivers over the speed limit at site } j \\ x_{6,j} &= \text{Traffic flow at site } j \\ x_{7,j} &= \text{Road classification} \end{aligned}$$

- This gives:

$$y_j \;=\; \exp\left\{2.08 - 0.04x_{2,j} - 0.01x_{4,j} + 0.00005x_{6,j} + 0.29x_{7,j}\right\}.$$

- The regression also gives an estimate of the "negative binomial over–dispersion parameter θ, $\hat{\theta} = 2.4496$, giving $\hat{gamma} = 1/2.4496 = 0.4082$ (common to all sites).

- We then use this PAM on each of the predictor variables at the *treated* sites $j = 1, \ldots, 67$ to get an estimate of casualty frequency at these sites.

4.1.2 Step 2: Bayesian bit

Plug the estimates of γ and μ_j into Equation (1) to get the EB estimate of casualty frequency at each treated site j.

4.1.3 Results

```
> cbind(before,mu,alpha,EB,after)
           before        mu      alpha        EB       after
 [1,]          20 1.4528615 0.21934974 15.931690           0
 [2,]           4 1.4301869 0.22205515  3.429360           0
 [3,]           9 0.7616679 0.34894493  6.125276           5
 [4,]           3 2.8195703 0.12647311  2.977180           0
 [5,]          17 1.5154547 0.21221250 13.713986           8
 [6,]           1 1.2069478 0.25274612  1.052305           1
 [7,]           3 0.7715166 0.34603190  2.228874           1
 [8,]           9 1.5856500 0.20474148  7.481975          17
 [9,]           4 1.4960553 0.21437436  3.463218           0
[10,]           7 5.3506604 0.07088691  6.883083           6
[11,]           6 2.7851340 0.12783695  5.589021           9
[12,]           8 2.7343999 0.12990073  7.315995           4
[13,]           5 1.3567552 0.23129368  4.157341           2
[14,]          12 1.7130921 0.19244127 10.020374           2
[15,]           3 4.4512992 0.08400606  3.121918           1
[16,]           4 2.3574299 0.14760670  3.757546           5
[17,]           8 3.3311436 0.10917067  7.490298           3
[18,]           3 3.4194920 0.10665088  3.044739          11
[19,]           6 2.8047886 0.12705495  5.594033           8
[20,]          11 3.5799384 0.10236025 10.240481          10
[21,]           7 5.4674430 0.06947799  6.893521          12
[22,]          21 1.1586491 0.26053698 15.830594          13
[23,]           4 1.1928128 0.25497752  3.284230           8
[24,]           1 2.6559789 0.13322523  1.220618           2
[25,]           3 1.6887009 0.19467973  2.744717           6
[26,]           4 3.6158543 0.10144666  3.961030           2
[27,]           3 2.8040033 0.12708601  2.975092           7
[28,]          11 1.6996572 0.19366783  9.198823           7
[29,]           5 1.8544082 0.18042210  4.432466           2
[30,]           8 5.8022731 0.06573218  7.855539           6
[31,]           3 3.8063530 0.09686128  3.078104          15
[32,]           9 1.6053091 0.20274249  7.500782          10
[33,]           7 1.7761689 0.18688434  6.023748           8
[34,]           3 1.6766983 0.19580047  2.740897           9
[35,]          16 2.5671126 0.13720434 14.156949           4
[36,]          13 4.5185546 0.08285930 12.297233          19
[37,]           8 2.5729949 0.13693362  7.256861           4
[38,]           4 3.1536509 0.11461077  3.902999           3
[39,]           4 2.3653752 0.14718386  3.759410           2
[40,]          28 3.8342926 0.09622339 25.674694          16
[41,]          13 3.5549846 0.10300475 12.027119          10
[42,]           2 4.6679851 0.08042014  2.214560           7
[43,]          15 3.4147638 0.10678279 13.762896          14
[44,]           1 1.4955738 0.21442858  1.106265           3
[45,]           4 3.7776263 0.09752603  3.978313           3
[46,]           7 1.3079474 0.23787164  5.646022           2
```

```
[47,]           8 2.0862039 0.16365635  7.032170          4
[48,]           4 4.9490367 0.07620116  4.072318         20
[49,]           7 2.9502425 0.12155226  6.507743          5
[50,]           5 2.6995561 0.13135715  4.697820          4
[51,]           9 2.8570545 0.12502124  8.232001         15
[52,]           6 8.4277103 0.04620107  6.112163          7
[53,]           7 5.0919986 0.07422054  6.858387          7
[54,]           2 2.9275374 0.12237961  2.113512          5
[55,]           8 3.9459665 0.09375551  7.619912          5
[56,]          16 7.8412636 0.04948545 15.596261          5
[57,]           7 2.8234310 0.12632202  6.472407          4
[58,]          11 2.2251190 0.15502310  9.639691          3
[59,]           8 1.0031321 0.28924536  5.976188          3
[60,]          11 2.7592477 0.12888170  9.937918         16
[61,]           4 1.8056743 0.18439367  3.595380          4
[62,]           5 3.9830888 0.09296294  4.905465          4
[63,]          12 3.9350031 0.09399217 11.241953         14
[64,]           8 1.7876912 0.18590372  6.845109         18
[65,]           7 6.2053084 0.06172640  6.950947          9
[66,]           6 1.6154651 0.20172502  5.115530         17
[67,]           7 2.8157738 0.12662203  6.470185          1
```

Total before	EB	Total after
505	487	457

Raw change: –**48**; After RTM: –**30**, so 18 would not have happened anyway.

4.2 Full Bayes analysis 1: completely analogous to EB

4.2.1 Model formulation

$$\begin{aligned} y_j &\sim \text{Poisson}(m_j), \\ m_j &\sim \text{Gamma}(\gamma, \gamma/\mu_j) \end{aligned}$$

at each *treated* site j. Then

$$\mu_j = \exp\left\{\beta_0 + \beta_2 x_{2,j} + \beta_4 x_{4,j} + \beta_6 x_{6,j} + \beta_7 x_{7,j}\right\},$$

with negative binomial error structure to account for over–dispersion (via dispersion parameter θ). We use independent priors here:

$$\beta_i \sim N(a_i, b_i^2), \qquad i = 0, 2, 4, 6, 7,$$

where a_i are set at the least squares estimates of each β_i and the b_i are large. We also use

$$\log\theta \sim N(0, 1000).$$

We use MCMC to (approximately) sample from the posterior. Why do this over EB? Overall results should be the same, but:

- We now have the full posterior distribution for $m_j|y_j$, not just the mean, so we can summarise the posterior in whichever we deem appropriate (.e.g. median, IQR).
- Estimates will have more realistic "standard errors" (posterior standard deviations here) since the regression parameters, and hence the μ_j and γ, are themselves allowed to vary.
- More 'natural'? All carried out in a single analysis?

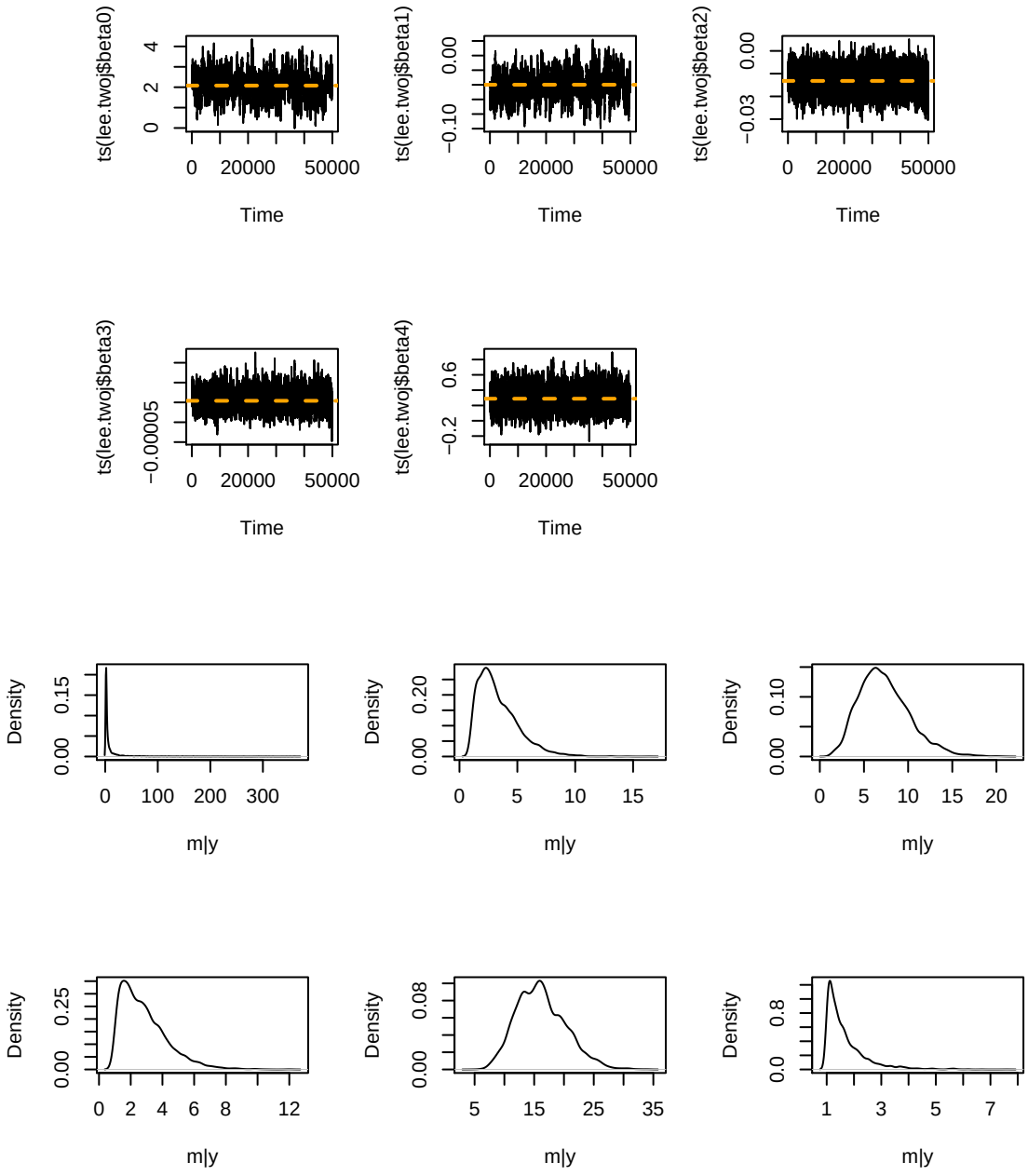
4.2.2 Some results

```
> cbind(before,FB,after)
           before Posterior mean   after
 [1,]          20      14.092974       0
 [2,]           4       3.350719       0
 [3,]           9       7.324299       5
 [4,]           3       2.849805       0
 [5,]          17      16.183616       8
 [6,]           1       1.666786       1
 [7,]           3       2.309379       1
 [8,]           9       8.333776      17
 [9,]           4       3.425840       0
[10,]           7       6.740952       6
[11,]           6       5.639401       9
[12,]           8       7.586068       4
[13,]           5       4.246250       2
[14,]          12      11.328424       2
[15,]           3       3.027134       1
[16,]           4       3.621659       5
[17,]           8       7.609435       3
[18,]           3       2.929134      11
[19,]           6       5.618162       8
[20,]          11      10.727340      10
[21,]           7       6.737679      12
[22,]          21      13.101511      13
[23,]           4       3.213070       8
[24,]           1       1.864663       2
[25,]           3       2.754231       6
[26,]           4       3.810000       2
[27,]           3       2.885068       7
[28,]          11      10.465991       7
[29,]           5       4.459143       2
[30,]           8       7.774054       6
[31,]           3       2.942385      15
[32,]           9       8.329543      10
[33,]           7       6.383752       8
[34,]           3       2.726399       9
[35,]          16      15.690655       4
[36,]          13      12.622610      19
[37,]           8       7.569486       4
[38,]           4       3.733292       3
[39,]           4       3.621979       2
[40,]          28      11.816646      16
[41,]          13      12.697167      10
[42,]           2       2.343542       7
[43,]          15      14.629688      14
[44,]           1       1.744495       3
[45,]           4       3.847725       3
[46,]           7       6.125552       2
[47,]           8       7.485812       4
[48,]           4       3.841488      20
[49,]           7       6.665166       5
[50,]           5       4.621435       4
[51,]           9       8.535517      15
[52,]           6       5.773673       7
[53,]           7       6.866117       7
[54,]           2       2.256728       5
[55,]           8       7.691306       5
[56,]          16      15.941782       5
[57,]           7       6.617685       4
[58,]          11      10.557913       3
[59,]           8       6.982950       3
[60,]          11      10.547301      16
[61,]           4       3.433688       4
[62,]           5       4.749919       4
[63,]          12      11.688079      14
[64,]           8       7.441437      18
[65,]           7       6.822525       9
[66,]           6       5.348946      17
[67,]           7       6.641251       1
```

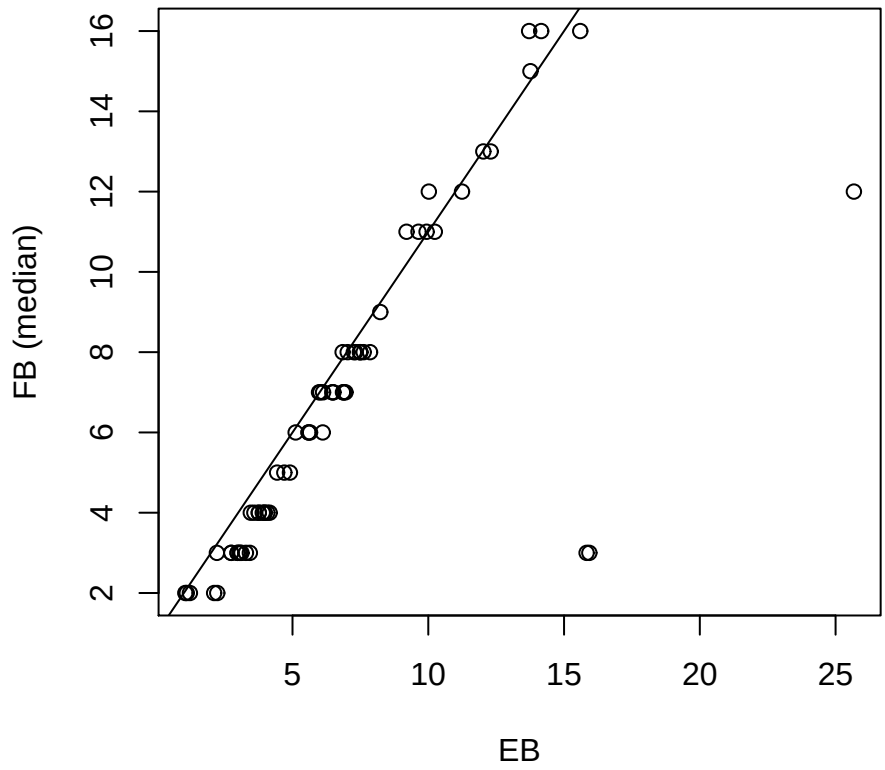
Giving

Total before	FB (mean	Total after
505	481	457

Raw change: −**48**; After RTM (using mean): −**24**. So 24 would not have happened anyway.



Using the posterior median gives:



5 Next steps

- Compare posterior summaries of spread to EB estimates of standard error.
- Investigate the use of other priors (c.f. Maher, 2010).
- Build in a proper trend term.
- Link to health costing.
- Build in random effects for sites?
- Re–visit zero–inflated models for fatals?