GETTING VALUE FOR MONEY FROM INVESTMENT IN ROAD SAFETY: ARE WE EVALUATING OUR SCHEMES CORRECTLY?

Peter Slater
North East Regional Road Safety Resource and Northumbria Safer Roads Initiative
Neil Thorpe
School of Civil Engineering and Geosciences, Newcastle University
Lee Fawcett and Joe Matthews
School of Mathematics & Statistics, Newcastle University

1. INTRODUCTION

1.1 Context for the research

One of the key issues when evaluating the costs and benefits of road safety interventions is the separation of scheme-effects from non-scheme effects to form an accurate picture of the 'true' impact of a measure on reducing collisions and casualties. Non-scheme effects that can blur the impact of a scheme come most notably from general trends in road safety statistics and Regression-To-Mean (RTM). The conventional approach to accounting for RTM is to adopt an Empirical Bayes statistical framework. Previous research by the authors on mobile safety cameras has suggested that this approach can under-estimate RTM effects thus inflating the apparent effectiveness of a road safety scheme leading to an over-optimistic impression of the return on the initial investment. The authors recommend that a Fully Bayesian approach is adopted instead, which accounts for RTM in a more realistic way and has the flexibility for trend effects to also be included.

This paper will therefore review the main arguments for preferring a Fully Bayesian framework over Empirical Bayes, describe the methodology for applying each method to road casualty data and then compare results from the application of each method to road casualty data collected at mobile safety camera sites operated by the Northumbria Safer Roads Initiative (NSRI) in the north east of England. These findings will also be compared to the results from the analysis of camera performance in the period immediately after the camera partnership was established to provide a long-term picture (2004-2014) of the 'true' impact of mobile safety cameras on casualty reduction in the region. The discussion will also focus on how the Fully Bayesian approach may be applied to road safety interventions in general to provide a more accurate assessment of their actual value for money.

2. EVALUATING THE BENEFITS OF ROAD SAFETY SCHEMES

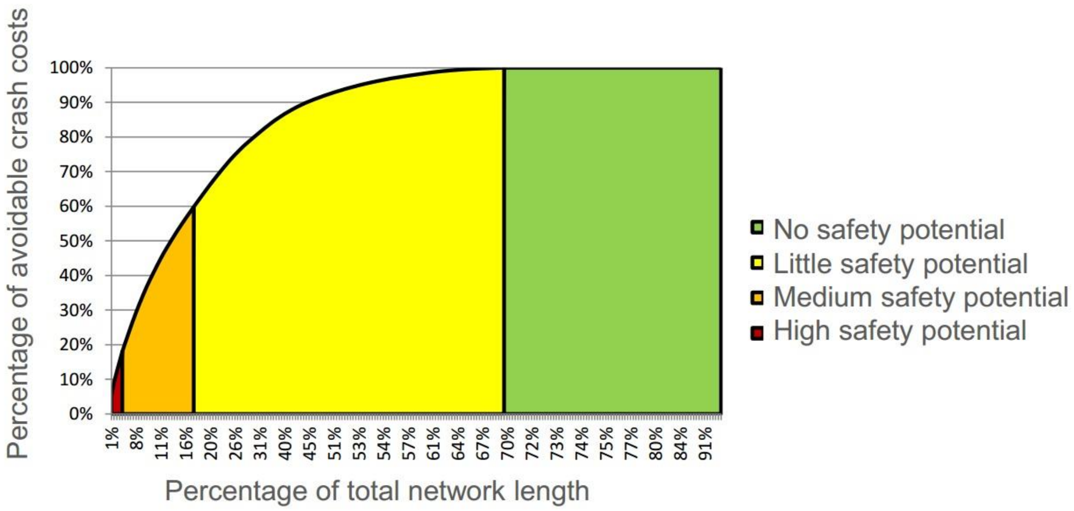
The Global Plan for the Decade of Action for Road Safety (2011-2020) developed by the World Health Organisation (WHO)identifies five key pillars to manage safety on road networks. These five pillars are:

- Pillar 1: building road safety management capacity;

- Pillar 2: improving the safety of road infrastructure and wider transport networks;
- Pillar 3: improving vehicle safety;
- Pillar 4: improving road user behaviour; and
- Pillar 5: improvementsto post-crash care (WHO, 2011).

Activity 6 of Pillar 1 identifies the need "to establish and support data systems for on-going monitoring and evaluation". Two of the principal functions for a road safety engineer to undertake are (1) to monitor accidents across the network to identify collision hotspots and (2) to evaluate the effectiveness of road safety interventions at reducing casualty numbers and casualty severity. These two functions relate closely to the efficient use of resources both in terms of ensuring that they are deployed at the right locations (the hotspots) and also that the deployment provides value-for-money (the evaluation). Each road traffic collision has a 'cost' associated to it. Collision hotspots are sites where the costs associated with the collisions are greater than we would expect for that location. In these cases, the potential exists to reduce collision costs through some form of road safety intervention – simply, where the difference between observed and expected costs is large then the potential savings are large, and where the difference is small, the potential savings are small. Evidence from Germany (see Figure 1 below) suggests that some 60% of avoidable collision costs can be attributed to a relatively small percentage of the network (17%).

Figure 1: Lorenz Curve for Network Safety Management



Source: Hoffman (2014)

Unfortunately the identification of collision hotspots across the network is made difficult by the issue of regression-to-mean (RTM), whereby a location may have an abnormally high collision rate during a monitoring period, due simply to it being at the top of a 'blip' with collisions subsequently returning to the underlying rate without any road safety intervention. Investment in these circumstances would not yield value-for-money as casualty reduction would have occurred anyway and would represent an inaccurate targeting of scarce resources. Similarly, RTM clouds the evaluation of road safety schemes – how much of an observed reduction in casualties is due to the treatment and how much due to other factors such as RTM and general trends in casualty rates? To help solve these problems associated with confounding factors, various statistical methods have been deployed to control for the effect of RTM both in the identification of hotspots to inform investment decisions and in the evaluation of road safety schemes to estimate value-for-money. These methods are discussed in the following section.

3. STATISTICAL METHOD

3.1 The standard approach: Empirical Bayes (EB)

The effect of RTM in the evaluation of road safety schemes is now well-known and well-documented (for example Hauer *et al*. (1988), Pendleton (1991), Mountain *et al*. (1992), Persaud and Dzbik (1993) and Kulmala (1994). The Empirical Bayes (EB) approach for accounting for the effects of RTM has become the 'gold standard' for road safety practitioners. This modelling approach aims to separate any change (usually a reduction) in casualty frequencies between a before and after period into various components, usually:

- general trends in casualty frequency;
- the amount by which we might have expected casualty frequencies to have changed anyway without the intervention (*i.e.*the RTM effect);and
- genuine treatment effect, that is, the change in casualty frequencies that we attribute to the safety intervention itself.

This is seen as an important component of any credible and realistic evaluation of a road safety intervention. Usually, a road safety measure is implemented at a site when casualty counts are abnormally high (in other words, we might be at the peak of a "blip" in terms of casualty frequencies). A simple comparison of before and after figures is bound to be biased, since we might expect some reduction in the after period anyway, even if nothing was done, simply because figures in the before period were abnormally high (relative to non-treated sites across the network).

Statistically, the standard EB method employed is rather simple, and can be thought of as the following four-stage process:

1. Identify a set of reference sites which are representative of the sites which have been treated with the safety measure – representative in terms of "predictor variables" which we suspect might impact upon casualty frequencies (*e.g.*traffic flow, road type and classification, speed limit, average observed speed) but *not* in terms of the observed casualty frequencies themselves (we would expect to see treated sites to have a more severe casualty profile than non-treated sites).

2. Use data from the reference sites to formulate a “casualty prediction model” (more commonly referred to as an APM after “accident prediction model”; see, for example, Mountain *et al*., 1992). This step involves the use of standard statistical modelling techniques widely available in basic software packages to construct an equation linking casualty frequencies to important predictor variables.
3. Estimate casualty frequencies at the treated sites using the APM constructed using data from the reference sites. The aim here is to obtain an estimate of what we would expect to see at the treated sites if we weren’t at the peak of a “blip”. This figure would be more in line with what we see at the reference sites.
4. Combine the APM estimate from step 3 with the observed casualty frequency in the before period at the treated sites to get the EB estimate of casualty frequency at each treated site. In other words, for each treated site two components are taken into account: the actual observed (and relatively high) casualty count from the before period, and the predicted value obtained from the APM. In fact, this EB estimate of casualty frequency is a weighted sum of these two components, and the value from the APM is given increasing weight as the estimated RTM effect increases. Conversely, as the RTM effect diminishes, so the observed casualty count is given more weight.

The difference between casualty counts in the before period and the EB estimate itself is usually recognised as the reduction we would expect to see anyway, even if no safety scheme had been employed – in other words, the reduction owing to RTM. The remaining reduction, that is, the difference between the EB estimate and observed counts from the after period, is attributed to the treatment itself. In reality, this estimated treatment effect is eroded further by other factors, most notably trend: crudely, trend is often accounted for by looking at historical casualty figures and applying a reduction factor based on observed trends from these figures.

3.2 Full Bayes: a more realistic approach to modelling

Although extremely useful, the standard EB approach for estimating RTM, as outlined in Section 3.1, was borne out of mathematical convenience. At the time (1980s – early 1990s), restrictive, and potentially unrealistic, statistical models were used to keep the mathematics analytically tractable. Since the early 1990s however, modern simulation-based procedures for performing complex statistical analyses have made the use of Bayesian statistics much more widespread and, in some instances, the norm. However, not until relatively recently have “Fully Bayesian” (FB) analyses made their way into the road safety literature (e.g. Miaou and Lord (2003); Li*et al.*, (2008); Maher and Mountain(2009)). In an FB analysis, we no longer need to worry about using statistical models which give full analytical solutions and can, instead, use model structures which more realistically capture the variability in our data. An FB analysis also gives more useful output. For example, the EB estimate of casualty frequency gives an estimate of the average casualty frequency at our treated sites, this average (by construction) being the mean. The results from an FB analysis give immediate access to the full distribution of estimated casualty frequencies at our treated sites, meaning other (perhaps more appropriate) summaries can be used. This full distribution also allows us to specify our uncertainties very easily, through standard deviations and credible intervals for our estimates. In fact, recent work by Fawcett and Thorpe (2013) has shown that an FB analysis gives a much more realistic assessment of this uncertainty than a standard EB analysis possibly can. More generally, as outlined and demonstrated by Fawcett and Thorpe (2013), working within an FB framework can allow much more complex, but potentially more realistic, models to be used – giving much more realistic assessments of road safety scheme effectiveness. Currently, the authors are working towards this alternative Fully Bayesian technique for road safety scheme evaluation being included in commercially availableaccident investigation software (such as PTV VISUM Safety), making such methods readily accessible to a large number of road safety practitioners.

3.3 Pre-analysis checks

Whether an EB or FB analysis is adopted, the methodology relies heavily on two assumptions:

1. the reference set of sites being representative of the treated sites, in terms of observations on the predictor variables; and
2. RTM being a valid assumption in the first place.

For example, if assumption 1 is violated, then it would not be valid to apply the APM, constructed using data at the reference sites, to sites receiving treatment. In such a circumstance, sites used in the reference set would have to be reconsidered and, perhaps, a new reference set obtained. Fawcett and Thorpe (2013) propose several pre-analysis checks for both assumptions, consisting of simple graphical diagnostics and more formal statistical checks. Again, it is the aim of the authors to have such diagnostics included in standard software used by practitioners.

4. RESULTS

4.1 Overview of the dataset

Across the Northumbria Police force area we have data on casualties, as well as several key predictor variables, for 51 sites treated with mobile safety cameras (the treated sites) and 46 sites not treated with any safety intervention at all (the reference sites). Specifically, we have casualty counts, as well as information on several predictor variables including: average speed, 85th percentile speed, average daily traffic flow, road classification, road type and speed limit. For the treated sites, we have well-defined “before” and “after periods”, each of nine years in duration (April 1994 – March 2003 for the before period, and April 2004 – March 2013 for the after period). Raw figures suggest a reduction of 508 casualties, in total across all treated sites, between the before and after periods – 168 of which are classified as "KSI" (killed or seriously injured).

4.2 Pre-analysis checks

The methods presented in Fawcett and Thorpe (2013) were implemented to check for compatibility between the reference and treated sites (assumption 1 in Section 3.3). In terms of the predictor variables, all diagnostic checks (both graphical and more formal statistical tests) revealed that the reference sites were representative of the treated sites. For example, both sets had a good mix of sites with different road classifications, speed limits and traffic flows. Of course, casualty counts at the treated sites were considerably higher than those at the reference sites.

From historical casualty records at the treated sites, we implemented simple methods from time series analysis (see for example Brockwell and Davis (2009)) to check the assumption of RTM and to estimate trend – assumption 2 in Section 3.3). This analysis revealed that the assumption of RTM is plausible, and that – since the mid-1970s – there has been a slight (but significant) negative trend in casualty frequencies generally, across all classes of severity (slight, serious and fatal).

4.2 RTM, trend and treatment effects

Various FB analyses, as outlined in Fawcett and Thorpe (2013), were employed to find the best-fitting statistical model. For this model, we estimate that of the total reduction in casualties of 508, we could have expected – on average – around 278 (55%) of these casualties not to have happened anyway due to RTM and trend. Thus, the remaining 230 (45%) might be attributed to the safety cameras themselves, notwithstanding any other effects that we have not accounted for (e.g. improved vehicle safety). The raw reduction of 168 KSI casualties can be divided into around 49 (29%), on average, that would not have happened anyway due to RTM and trend, and 119 (71%) being due to the safety cameras (again, not including any other possible effects). At a site level, there was considerable variation in the RTM/trend effect. For all severities of casualty combined, we see a reduction of between 23% and 79% owing to RTM and trend (c.f. 55% across all sites together); for KSI, at a site level, this reduction is between 20% and 37%.

In comparison, the EB estimate on the same data suggests that 259 casualties would have been saved as a result of the mobile safety cameras – 29 casualties more than suggested by the FB method implying that the mobile cameras have been more effective and the RTM effect having less of an impact.

4.3 Cost effectiveness

Following Thorpe and Fawcett (2012) and Fawcett and Thorpe (2013) we attempt to estimate the financial consequences of the implementation of the mobile safety cameras across the Northumbria region. We do this in two ways: (1) by estimating cost savings to local National Health Service (NHS) secondary healthcare providers via a large scale data linkage exercise, linking known police records of road traffic accidents to hospital records, and (2) by attempting to account for costs associated with lost output, pain grief and suffering via the Department for Transport's (DfT) valuation of road traffic accident casualties.

A large, multi-stage data linking exercise took place (Colligan*et al*. (2008)) to link police collision data in the Northumbria Police Force area to NHS patient data. This was attempted using unique identifiers which are collected by both the police, at the scene of the accident on STATS19 forms, and the hospital involved. NHS Accident and Emergency (A&E) patients fall into one of eight Health Resource Group (HRG) categories depending on the severity of their injuries and the extent of treatment required. These A&E HRGs range from high cost categories which include computerised tomography scans and magnetic resonance imaging scans, to relatively low cost categories, involving more routine urine/bacteriological investigations. Some A&E patients then require admission to hospital for in-patient treatment; in-patients are allocated to one of some 700 in-patient HRGs (see Colligan *et al*., (2008) for full details). Each A&E HRG has an associated financial tariff, as does each in-patient HRG. Overall individual in-patient tariffs are calculated as a function of time, and so instead of considering the 700 inpatient HRGs themselves we discretise the in-patient tariffs into groups of £500 (where £0 is used for A&E admissions not requiring in-patient treatment). Based on this data linkage exercise, we use the proportion of casualties falling into each A&E HRG/in-patient tariff category combination as the expected proportions that *would* have occurred if the mobile safety cameras had not been in place, in other words if the 230 casualties that were saved by the cameras had actually occurred. Across our 9 year after period, this results in estimated (average) total NHS savings of £140,000 when implemented within our Fully Bayesian analysis. For comparison, the traditional EB analysis gives corresponding savings of almost £160,000 – almost surely exaggerated by the way in which the EB analysis summarises our findings using the arithmetic mean as the key summary for RTM.

To account for costs incurred elsewhere, including pain, grief and suffering, we also attempt to estimate the cost of the 230 casualties prevented by the safety cameras using the DfT's figures relating to the valuation of casualties as a result of road traffic accidents. Currently, such a casualty is estimated to cost approximately £48,971 (this is an average figure, of course, and will vary depending on the severity of the casualty; see Department for Transport (2014)), thus resulting in an estimated (average) saving, across the 9 year after period and for our safety camera sites, of some £11.3 million. Again, for comparison the traditional EB analysis gives (exaggerated) corresponding savings of almost £12.7 million.

5. CONCLUSIONS

This paper has described how the job of a road safety engineer in identifying collision hotspots to inform investment decisions and the evaluation of road safety interventions to determine value-for-money is problematic due to the effect of

confounding factors such as regression-to-mean (RTM) and general trends in casualty rates. These confounding factors can lead to the inefficient targeting of resources and inaccurate estimates of the true impact of a road safety measure. The paper reviews the method currently adopted by many road safety engineers that is based on an Empirical Bayes statistical framework, and suggests that an alternative approach based on Full Bayes may be more flexible and provide more realistic estimates of the magnitude of RTM. The benefits of this include more accurate targeting of resources to 'real' collision hotspots and more realistic estimates of returns on investments. The two approaches have been demonstrated using casualty data from 51 mobile camera sites and 46 reference sites in Northumbria over a nine year before and nine year after period. The results from the EB method suggests that cameras have been more effective compared to the analysis using the FB. As we argue that the FB approach provides more realistic results, this suggests that the current 'gold standard' approach for accounting for RTM in road safety management using Empirical Bayes can under-estimate the effect of RTM and therefore over-estimate the impacts of site-based road safety measures.

References

Brockwell, P.J., Davis, R.A. (2009) *Time Series: Theory and Methods*, Springer-Verlag, New York.

Colligan, J., Goulbourne, L.E., Thorpe, N., Fawcett, L., and McNay, A. (2008)'Mobile Speed Cameras and the Demand forSecondary Health Care in Northumbria' Northumbria Safety Camera Partnership, Newcastle upon Tyne.

Department for Transport (2014) TAG Unit A4.1 Social Impact Appraisal https://www.gov.uk/transport-analysis-guidance-webtag

Fawcett, L., Thorpe, N. (2013) Mobile safety cameras: estimating casualty reductions and the demand for secondary healthcare. Journal of Applied Statistics 40(11), 2385-2406http://www.tandfonline.com/doi/full/10.1080/02664763.2013.817547

Hauer, E., Ng, J., and Lovell, J. (1988) Estimation of safety at signalised intersections, **Transportation Research Record** 1467, 30–37.

Hoffman, T. (2014) 'Road Safety – Network Safety Management with PTV VISUM Safety' Presentation to PTV Vision Traffic User Group Meeting: PTV Visum Safety, Berlin (May)

Kulmala, R. (1994) Measuring the safety effect of road accident measures at junctions, **Accident Analysis and Prevention** 6, 781–794.

Li, W., Carriquiry, A., Pawlovich, M., and Welch, T. (2008) The choice of statistical models in road safety ountermeasureeffectiveness studies in Iowa, **Accident Analysis and Prevention** 40, 1531–1542.

Maher, M.J., Mountain,L.J. (2009)The sensitivity of estimates of regression to the mean, **Accident Analysis and Prevention** 41, 861–868.

Miaou, S.P., Lord, D. (2003)Modelling traffic crash-flow relationships for intersections: Dispersion parameter, functionalform, and Bayes versus empirical Bayes methods, **Transportation Research Record** 1840, 1310–1340.

Mountain, L., Fawaz, B., and Sineng, L. (1992)The assessment of changes in accident frequencies on link segments: Acomparison of four methods, **Traffic Engineering and Control**, 33, 429–431.

Pendleton, O.J. (1991) Application of new accident analysis methodologies, Vol.1: New methodologies, US Department ofTransportation, Report FHWA-RD-90-091,Washington, DC.

Persaud, B., and Dzbik, L. (1993) Accident prediction models for freeways, **Transportation Research Record** 1401, 55–60.

Thorpe, N., Fawcett, L. (2012) Linking road casualty and clinical data to assess the effectiveness of mobile safety enforcement cameras: a before and after study BMJ Open 2012 2(6) e001304 http://bmjopen.bmj.com/content/2/6/e001304?ct

World Health Organisation (2011) ‘Global Planfor the Decade of Actionfor Road Safety 2011-2020’ WHO (Geneva) http://www.who.int/roadsafety/decade_of_action/plan/plan_english.pdf?ua=1