Use of Accident Prediction Models in Road Safety Management – An Irish case study

Nathan Harpham[1], Caroline Wallbank[1[0000-0003-0375-6922]], John Fletcher[1], Lynne Smith[1[0009-0007-9137-3724]], Ian D'Arcy[2], Dr. Suzanne Meade[3]

[1] TRL Limited. Crowthorne House, Nine Mile Ride, Wokingham, Berkshire. RG40 3GA, England
Corresponding author: `nharpham@trl.co.uk`
[2] Arup, 50 Ringsend Road, Dublin 4, D04 T6X0, Ireland
[3] Transport Infrastructure Ireland, Parkgate Street, Dublin 8, D08 DK10, Ireland

Abstract. Evaluation of road safety measures can be a challenging element of road safety management systems in Europe. To deliver Vision Zero and implement the Road Infrastructure Safety Management Directive, national road authorities need reliable estimation tools for road safety countermeasures. Accident Prediction Models (APMs) provide an objective and cost-effective way to analyse potential safety improvements and estimate the potential impact in terms of collision reduction. However, most National Road Administrations (NRAs) do not develop or use APMs. The objective of this paper is to present research undertaken for Ireland's first APM including the modelling technique used and the data challenges faced. The primary aim of the APM development is to provide local (Irish) estimates for Crash Modification Factors (CMFs) to feed into a tool for use by Road Safety Engineers when estimating the potential collision savings of various interventions.

Keywords: Accident Prediction Models, Crash Modification Factors, road safety, safe systems, collision data.

1 Introduction

The aim of this work was to develop Ireland's first Accident Prediction Model (APM) to provide Irish Crash Modification Factors (CMFs) for Transport Infrastructure Ireland (TII), local authorities and road safety practitioners to identify cost effective road safety interventions and measures to reduce road traffic collisions and achieve targets towards Vision Zero [1]. The objectives of the project are **(1)** understand the extent to which APMs can feasibly be developed from Irish data using established methodologies (identified through a literature review), and **(2)** to develop a tool for practitioners to make better use of APM findings and CMFs.

2 Feasibility of model development

2.1 Literature review

A review of international literature was carried out to identify applicable modelling approaches in the European context. The literature review also informed the expected factors contributing to collisions. Generalised linear models (GLM) as accident prediction models (APM), use typically either a Negative Binomial (NB) or Poisson (P) distribution error structure. APMs, of various functional forms, their purpose is twofold, to predict the frequency of accidents or attempt to explain the association between different accident types or severities and several independent variables [2].

More than 25 published papers and reports were reviewed to understand the types of APM development (e.g., the datasets, variables, and methods for assigning these to the network) and the statistical approaches used to develop the models. The key conclusions from the review were:
(1) Most papers reviewed used five or six years of collision data and modelled all injury collisions as this offered a good balance between obtaining sufficient collisions per road segment for the modelling and reducing the chance of there being major differences in the road factors present over time.
(2) Traffic flow is always a highly significant factor that explains collision occurrence in APMs [3].
(3) The approach applied to divide the roads into segments is very important. The aim is to define segments with relatively few zero collision counts whilst capturing enough variability in the other explanatory parameters. A recommended approach is to divide the network into segments with the same flows and/or other specific features, mainly curvature [4].
(4) Many APM have commonly occurring significant variables. (Variables that were significant in developed APMs were investigated within the Irish data, *see section 2.2*).
(5) Road type characteristics are significant. Motorway, dual carriageway and undivided roads will have differing parameters and characteristics. For this reason, models should be developed for specific road types.

The most common statistical approach for the development of APMs were GLMs; both P and NB GLMs were used in the literature. While both the P and NB models can take account of excess zeros in the data, or “overdispersion”, where significant numbers of segments have no (zero) collisions, zero-inflated models should be considered [2].

2.2 Data Collection

Pre-existing data sources for a range of road features and traffic characteristics on the road network from TII asset databases were assessed for use in the modelling. Throughout this process several challenges with using pre-existing data were identified and resolved. Collision records have co-ordinates which permit linking to road segments. Including damage only collisions in addition to injury incidents increases the total number of collisions for modelling, reducing the number of segments on which there are zero collisions (necessary for the modelling). In Ireland, collisions cannot be accurately linked to carriageways so both directions were combined for the modelling.

Based on the data assessment and exploratory analysis, decisions were made on how to divide the road into the short sections (‘segments’) that will be modelled. Segments were created for each of the road types according to variation in radius (curvature), traffic flow (AADT) and number of lanes. Sufficiently large changes in these parameters mark the boundary points between segments. To reduce the frequency of segments with a zero-collision count, a minimum segment length (200m) was imposed. A maximum segment length (5km) was also used.

Roundabouts, link roads and ramps have been excluded from the modelling as these combined accounted for less than 5% of the network length and 10% of collisions. The number of a major and minor junctions on each mainline was identified for each segment but, due to data availability, it has not been possible to model the risk at junctions separately from the mainline. More detailed information on junction types, including whether the junction is signalised or not, the number of turning movements, the number of arms and the flows on each of these would enable separate junction models to be developed.

Four models, one for each of the road types (motorways, dual carriageways, single carriageway and legacy roads) which make up the TII network, were deemed feasible to develop based on the data available and methodologies identified in the literature. Aggregation of the potential explanatory variables on the segments was carried out.

The following variables were tested for inclusion in the models:

- Traffic information (AADT, %HGVs)
- Road characteristics (gradient, crossfall, radius, friction coefficient, hard shoulder & median widths, verge & median barrier presence)
- Junction/access information (minor & major junction density, business, commercial & residential access density)
- Road classification (urban/rural, flag for segments on the M50 as this was considered a high-risk road, flag for 1+1 and 2+1 lane segments)

No suitable dataset for vehicle speeds was identified during the assessment and as a result, this was excluded from the list above. For some variables, additional data collection was needed in the form of a visual inspection using Google Earth imagery. As a result, these variables (hard shoulder and median widths, verge barriers) were only used some of the models as the scale of data collection needed for the whole network was outside of the scope of this project. TII are investing in more reliable data collection methods which could be used for future modelling activities.

3 Model results

3.1 Modelling process

NB zero-inflated models were determined to be the best fit for the data. Zero inflated models involve two different model processes:

1. A binomial logit model to model whether the observation is zero or not. This model represents the "structural/excessive zeros" – these are observations which are always zero.
2. A NB model to model the non-zero count data. Within this model, "sampling zeros" are modelled for those segments which are exposed to the risk but do not report the outcome (collisions) during the data.

For the first part of the modelling (the logit model), AADT was the only predictor included. Whilst this variable was not significant in all of the road type models, it makes logical sense that as the AADT increases, the probability of no collisions on a segment decreases – this is supported by the direction of the sign (negative) for the coefficient in each of the four models. For each road type, a base NB model was developed using AADT and segment length which are known from the literature to be the key variables affecting collision risk. For the other variables a stepwise variable selection process, according to how significant variables were (using their p-values) and the change in Akaike's Information Criterion (AIC) was used for variable selection. For each of the final models selected, various goodness of fit measures were used to compare the base model to the final model. Predictive performance of the model was assessed using K-fold cross-validation.

3.2 Model results

The results from the NB part of the zero-inflated model for the motorway network are illustrated in **Table 1** below. Models for the other three road types were also created.

Table 1. Variables included in the Negative Binomial motorway model

Variable	Coefficient	p-value
Intercept	-9.192	p<0.001
Log(segment length)	0.765	p<0.001
Log(AADT)	1.157	p<0.001
Gradient	0.176	p<0.001
HGV %	1.804	p<0.001
Radius	-0.187	0.004

The direction of the coefficients (positive or negative) is as expected. The segment length or AADT have positive coefficients which indicate that as segment length and AADT increase, collision risk increases:

i) The segment length coefficient is less than 1 which indicates that as you double segment length, you don't quite double collision risk.

ii) The AADT coefficient is slightly greater than 1, suggesting that as the flow increases, the collision risk increases.

Increases in gradient and increases in the percentage of vehicles which are HGVs, both increase collision risk, possibly due to increased speed differentials between vehicle types. The results also show that as the radius increases (i.e., the road becomes less bendy), the collision risk decreases. Model predictions compared with the actual collisions on each segment, **Fig. 1** below, are relatively close to y=x the line.

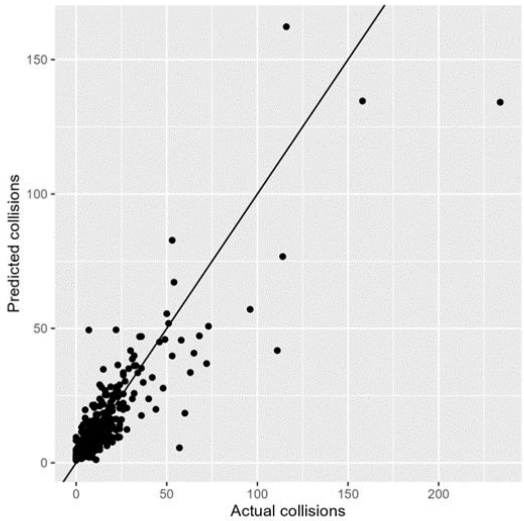

Fig. 1. Model predictions compared to actual collisions by segment for the motorway model.

The results from all four NB models indicate that collisions decrease as the road environmental conditions improved. For example, reducing the number of accesses or junctions, improved pavement condition, and increasing the proportion of dual carriageway with a median barrier, all decreased collision risk. These results agree with findings in other jurisdictions and demonstrate that these models can be useful for the evaluation of safety countermeasures at the local level, supporting Ireland's Vision Zero aims.

4 CMFs for Road Safety Practitioners

The previous section answered the first objective of the research and informed the second objective. After an APM has been developed, any modification to the predictions from the model must account for geometric design or traffic differences between the base conditions of the model and the conditions of the site being considered. This is determined using CMFs. This provides a means to predict how system changes will impact safety in terms of collision/casualty occurrence [5]. Road safety practitioners require this information for the economic appraisal of countermeasures and to estimate potential collision reduction. While it was possible to develop APM for Irish data, there were limitations to the scope of the CMFs due to the availability of data. Therefore, CMFs developed from other APMs were integrated into a tool that could provide a wide scope of potential CMFs for practitioners by integrating data from a CMF database [6].

The tool facilitates use of the new APM and other CMFs for multiple countermeasures to compare the impact of different combinations of safety measures on collisions. This tool provides a practical means for a wide cohort of users that would not typically use CMFs or APMs. The tool can be obtained on the TII Publications website [7].

5 Conclusions

Ireland is a relatively small country with small road collision numbers relative to other countries. This can present methodological challenges when using GLMs to develop APMs. This research demonstrates that APMs can be developed under these circumstances, using a zero-inflated modelling approach, to provide empirical evidence-based information for road safety programmes. APMs are a more cost-efficient way to carry out evaluation of safety interventions compared to studies of individual features or sites. This research highlighted significant limitations associated with the use of existing datasets collected for other purposes. Further work should be considered to include suitable speed data might be collected in the future; as a minimum this should include mean and 85th percentile speeds to enable the range of operating speeds to be understood. While asset data is routinely collected by national road authorities, these data may not be easily transferred to other purposes; authorities should consider data analytics and digitisation for all data collection as part of a Safe Systems to achieve Vision Zero and implement the Road Infrastructure Safety Management Directive.

References

1. Government of Ireland, (2021). Our Journey Towards Vision Zero Ireland’s Government Road Safety Strategy 2021–2030, Road Safety Authority, Ballina. https://www.rsa.ie/about/safety-strategy-2021-2030
2. Lord, D. and Mannering, F., (2010). *The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives*. Transportation Research Part A: Policy Pract., 44, 5, 291–305. https://ceprofs.civil.tamu.edu/dlord/Papers/Lord-Mannering_Review.pdf
3. Elvik, R., Høye, A., Vaa, T., & & Sørensen, M. (2009). The Handbook of Road Safety Measures. Oslo: The Emerald Group Publishing Limited.
4. Turner, S., Singh, R., & Nates, G. (2012). The next generation of rural road crash prediction models: final report. Wellington: NZ Transport Agency.
5. Yannis, G., Dragomanovits, A., Laiou, A., Richter, T., Ruhl, S., La Torre, F., Li, H. (2016). *Use of accident prediction models in road safety management – an international inquiry*. Transport Research Procedia, 4257-4266.
6. CMF Clearinghouse, https://www.cmfclearinghouse.org, last accessed September 2023.
7. TII Publications, www.tiipublications.ie.