THE FUTURE OF TRANSPORT

PUBLISHED PROJECT REPORT PPR2031

Collision Prediction Model for the Irish National Road Network

Phase 2 Report

C Wallbank, N Harpham & J Fletcher

Report details

Report prepared for:	Transport Ireland Infrastructure, Suzanne Meade		
Project/customer reference:	TII268 Collision Prediction Modelling		
Copyright:	© TRL Limited		
Report date:	November 2023		
Report status/version:	Final - version 2.0		
Quality approval:			
Warsame Mohamed (Project Manager)		Lynne Smith (Technical Reviewer)	

Disclaimer

This report has been produced by TRL Limited (TRL) under a contract with Transport Ireland Infrastructure. Any views expressed in this report are not necessarily those of Transport Ireland Infrastructure.

The information contained herein is the property of TRL Limited and does not necessarily reflect the views or policies of the customer for whom this report was prepared. Whilst every effort has been made to ensure that the matter presented in this report is relevant, accurate and up-to-date, TRL Limited cannot accept any liability for any error or omission, or reliance on part or all of the content in another context.

When purchased in hard copy, this publication is printed on paper that is FSC (Forest Stewardship Council) and TCF (Totally Chlorine Free) registered.

Executive Summary

This report sets out the final methods, findings, outputs and conclusions for the project: 'TII268 Lot1 Collision Prediction Model for the Irish National Road Network'. The aim of the project was to develop Accident Predictive Models (APMs) using pre-existing data for TII's road network. APMs relate crash numbers to physical road features and exposure (traffic) data. Regression techniques were used to provide quantitative estimates of the impact of these road characteristics on safety. Although the relationships developed are not strictly causal, this approach is a well-established and practical way to understand how road features perform with respect to safety in the field, locally. Practitioners can use the developed safety effectiveness estimates (or Crash Modification Factors (CMFs)) to assist them to identify the most appropriate road designs in given circumstances. These can also be used in economic appraisal for road design upgrades over the long term; this again assists decision making and improves transparency.

Phase 1

The Phase 1 client report (Chowdhury, et al., 2022) gave details of the preparatory review phases of the project: this assessed the quality of available data and reviewed potential statistical approaches. It was acknowledged that using existing data rather than undertaking bespoke, targeted data collection is challenging. Also, as a small country with generally low traffic flows, Ireland has a relatively low density of road collisions. However, this does not mean the network is low risk or safe. Small sample numbers of crashes per modelled road section also pose difficulties for the statistical methods used to develop the APMs.

Phase 2 Modelling

The road was segmented into homogeneous sections and the crash occurrence was modelled using information on the road features across these sections. Roads were sectioned on the basis of consistent flow, bendiness and number of lanes, in line with the recommendations from the Phase 1 findings. In Phase 2, threshold values for these features were chosen to reduce the number of sections with zero crash count. All crash severities (including damage crashes) were included to maximise the crash count per modelled section.

A range of modelling approaches were tested to identify the approach that gave the best models. The zero-inflated approach, using a Negative binomial distribution, was found to produce the best fit between crashes and explanatory variables. This approach is specifically applied when there are a high proportion of modelled sections with zero crash counts. Models for each road type were obtained, the fit was best on the motorways (where flows and crash density were highest).

Practical models results

Some of the key findings from the models were:

- Reducing the number, or improving the safety of, minor junctions and access points onto the network could reduce collision risk.
- On dual carriageways, increasing the proportion of median barriers decreases the risk on a segment.

- It is important to ensure the skid resistance (CSC %) meets the defined minimum thresholds on single and legacy roads.
- The geometry of the road influences collision risk: gradient and radius were common significant predictors of collision risk across all models.

Tool planning

This work was conducted to provide local (Irish) estimates of the safety performance of road features. The model outputs were fed into a tool to support engineering safety practitioners to select effective infrastructure interventions (see Figure 1 below).

Practitioners were asked to complete a survey to understand their needs for the tool. Following this, two workshops were undertaken to capture the detailed needs and challenges that road safety staff currently face. Amongst the findings were that:

- Ways to avoid double counting crash savings of measures when used in combination were needed.
- First Year Rate of Return was the preferred economic appraisal approach.
- Other specific engineering countermeasures (those not modelled) needed to be available in the tool and their effectiveness should be obtained from the 'Clearinghouse' source; this should be limited to the best quality study findings.

These and other captured aspects were then fed into the tool design as shown in Figure 1. The tool was developed in MS Excel and can be downloaded from the TII website.

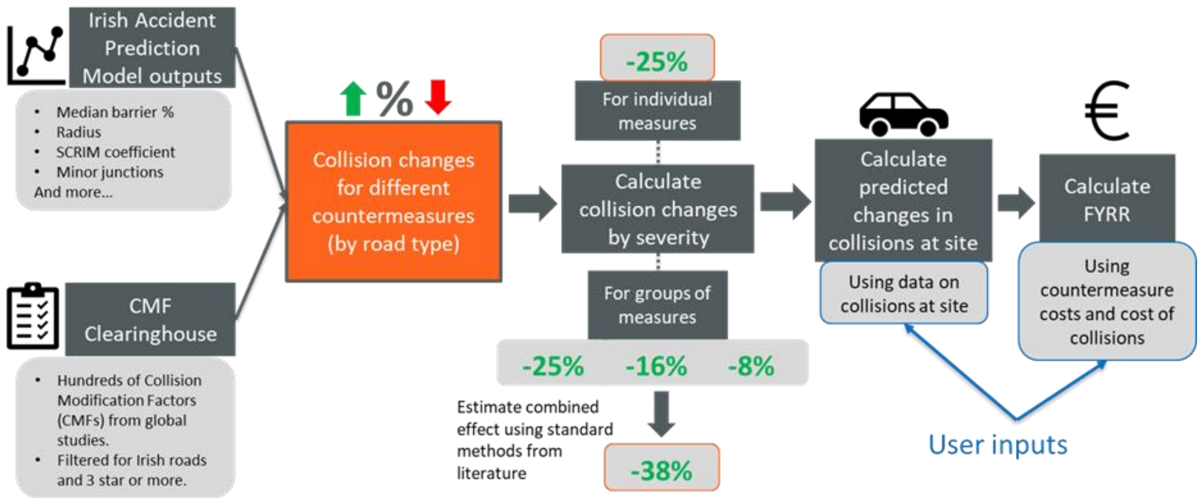


Figure 1: Summary of the process flow for the Collision Reduction Calculator

Future modelling recommendations

Data was not available for some important explanatory variables in the modelling. As a result, the coefficients presented in this report should be treated as indicative and the models should be updated as new data sources are available. Relating to this issue, availability of the following data types would benefit the model fit and robustness:

1. Robust data on traffic speeds.
2. Flow data by road user type (e.g. motorcycles, pedestrians, pedal cycles).
3. Presence of rumble strips, street lighting, pedestrian crossing facilities and cycle facilities.
4. Separate junction models should be developed; this would require substantially more detailed data to be collected.
5. Accurate assignment of collisions to individual carriageways and specific roads (e.g. where one road crosses over the top of another) would also improve the quality of the data.

Future potential improvements to the tool

The project identified that the tool could be extended and improved by the following changes:

- Countermeasure effectiveness broken down by the different collision sub-types targeted by the specific measure.
- The user could be permitted to specify the amount of overlap between countermeasures - to give more accurate combined reductions that account for the potential double counting of benefits.
- Links to other data sources for benefit estimations, beyond the Clearinghouse, could be beneficial.
- Countermeasures could be categorised in the tool by relevance or applicability to Irish roads.

Table of Contents

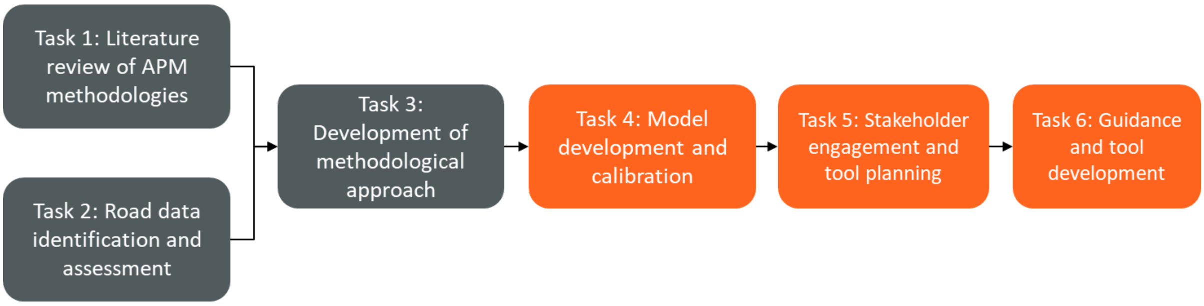
1 Purpose of this project

The aim of this work was to develop Ireland's first Accident Prediction Model (APM) and to use this to provide Irish Crash Modification Factors (CMFs). These locally derived CMFs are to benefit road safety practitioners at Transport Infrastructure Ireland (TII) and local authorities as they quantify the effective road safety interventions. They are used to identify the measures which will be most appropriate to reduce road traffic collisions and casualties. CMFs are also important for the economic appraisal of countermeasures. These can therefore help staff to programme targeted, cost-effective and proactive interventions.

The project investigated:

- The extent to which APMs can feasibly be developed from available Irish data sources.
- How the APMs developed and CMFs from other sources can be used to develop a decision tool for practitioners to provide effective information to inform road safety decisions into the future.

There were six tasks to address these aims:



The findings, conclusions and recommendations from the first three tasks are documented in the interim report (Chowdhury, et al., 2022). This report assessed the feasibility of developing robust APMs for the Irish national road network. It made recommendations on the best approach based on the methodological review findings and the availability and quality of data.

This report covers the findings from the last three tasks. Section 2 presents and overview of the model development including the data used, road segmentation approach, modelling methodology and results. It also makes some recommendations which should improve future modelling. Section 3 details the planning process for the tool (Collision Reduction Calculator), which involved a survey of, and workshops with, Road Safety Engineers to understand their needs. Section 4 details the process flow for the calculator, the Irish CMFs included and makes some suggestions for future improvements to the calculator.

The calculator is available to download from the TII Publications website: https://www.tiipublications.ie/.

2 Model development

2.1 Overview of methodological approach

As outlined in the interim report (Chowdhury, et al., 2022), the APMs were developed using statistical models with a negative binomial distribution, as this was assessed as the most appropriate for the response variable (number of collisions (including damage only collisions) on each section).

Four separate APMs were developed covering the following road types (see Figure 2):

- Mainline: motorway.
- Mainline: dual carriageway.
- Non-legacy road mainline: single carriageway.
- Legacy roads (subnet 3 and 4).

For each road type, the network was first segmented into homogenous sections (see Section 2.2) and then the APMs developed using these segments (see Section 2.3).

Note that 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 major and minor junctions was identified for each mainline segment but, due to data availability, it has not been possible to model the risk at specific junctions separately from the mainline.

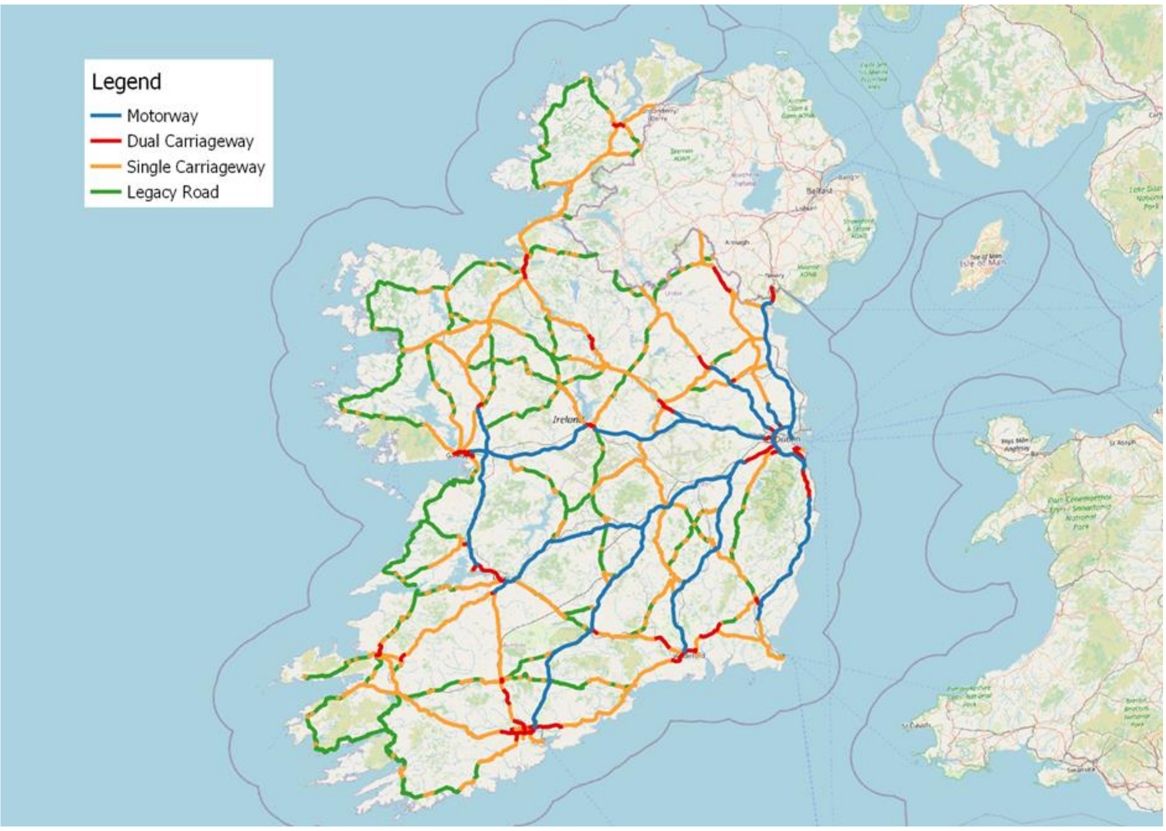


Figure 2: Road Types

2.2 Segmenting the network

This Section describes the process of segmenting the network for modelling across the four different road types. Section 2.2.1 outlines the data sources used; Section 2.2.2 outlines the method and results of the segmentation.

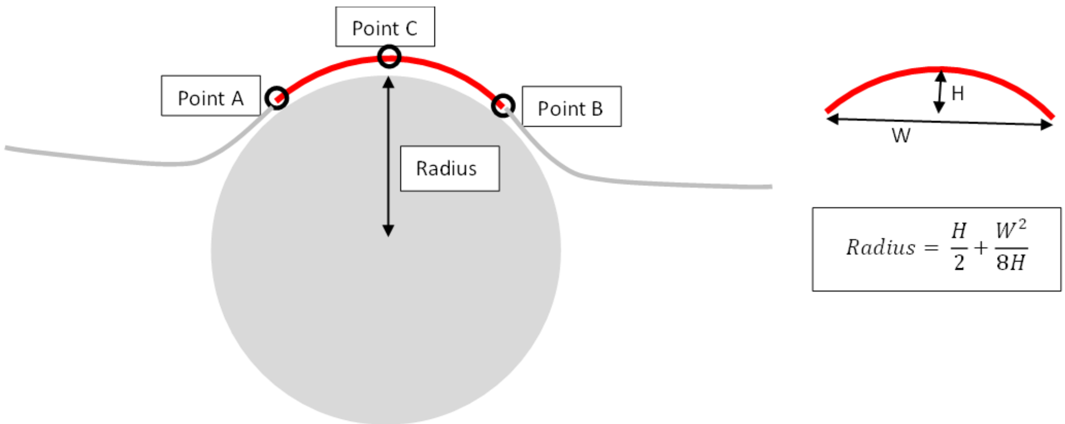
2.2.1 Data sources

As set out in the interim report (Chowdhury, et al., 2022), it was intended that the TII GIS base data would be used as the reference layer to link other datasets. However, after extensive further investigation, some issues were identified with these data (namely with the junction counts/types and some small discrepancies in the road type). As a result, a base layer was created in two steps:

1. The PRIME2 Q1 2020 database (chosen to align to the end of the collision data period) was used to classify the network into motorway/dual/single.
2. This classification was combined with the TII GIS base layer data to distinguish the single carriageway into single (non-legacy) and legacy networks. This also ensured that only National roads were included in the modelling.

Segments were created for each of the road types according to variation in radius (**curvature**), traffic flow (**AADT –** Average Annual Daily Traffic) and **number of lanes**. Where sufficiently large changes in these parameters occur, boundary points between segments were created.

The radius variable in the Pavement Management Survey (PMS) data was used for the motorway and dual carriageway segmentation. For single carriageway and legacy roads, an alternative method was used to generate radius values for segmenting, as the PMS data was not sufficiently accurate. The road network was split into 250m sections and radius values were calculated for each section using the GPS co-ordinates A, B and C - the start, end, and midpoint of each section (see Figure 3). A standard mathematical formula gives the radius of the circle containing these 3 points – which is assigned to point C. Other section lengths were tested, such as 100m and 500m, but these resulted in less suitable segments for modelling.



$$Radius = \frac{H}{2} + \frac{W^2}{8H}$$

Figure 3: Alternative method for calculating radius values on the road network

Adjustments were needed to the calculated radius values for single and legacy roads to ensure these were similar in scale to those from the PMS survey data. This involved scaling the calculated values and applying a maximum value (since very straight roads would have an infinite radius in this calculation). This means that the radius values for motorways and dual carriageways are not directly comparable to those for single and legacy roads. There may therefore be some differences in the model coefficients for this variable across road types.

The AADT variable was taken from the 2015 to 2019 traffic data – averaging across the five years to match the collision period. The AADT for light vehicles and heavy vehicles were summed for each section to give an overall AADT for segmenting. The 'number of lanes' variable was taken from the traffic data. Some 3-lane sections of the network were incorrectly labelled as 2-lane, so manual adjustments were made to improve the accuracy of this dataset before segmenting.

Table 1 summarises the variables and source of these data for the segmentation.

Table 1: Variables used to segment the network

Variable	Data source
Traffic flow (AADT)	Traffic data from 2015 to 2019
Radius (curvature)	PMS survey data - motorways and dual carriageways Alternative method - single carriageways and legacy roads
Number of lanes	Traffic data with manual corrections

Once the segments were created, data from other sources were linked to each of these; the initial list of variables was taken to be those identified in the phase 1 report (Table 12 (Chowdhury, et al., 2022)). Additional variables have also since been incorporated, and improvements made to some:

- Width of the hard shoulder and median (motorways only) – measurements at the 1km level were obtained using Google Earth imagery, averaged across each segment and then categorised as in Table 8.
- Road hazard (risk rating) data – data provided by TII classifying the risk at different points on the network as 'high', 'medium' or 'low' was considered for inclusion.
- Location of 2+1 and 1+1 roads – a yes/no marker for each of these road types in the data. Each segment is entirely 2+1 or 1+1, or neither. The 2+1 and 1+1 segments were included in the single carriageway dataset for comparison with other 2 and 3-lane segments.
- Presence of a verge barrier (motorways and dual carriageways) and median barrier (dual carriageways only) - whilst these data are collected in the Vehicle Restraint System (VRS) dataset, further examination showed that there were substantial amounts of missing data. In locations where there were no barriers recorded, a manual inspection of Google Earth imagery was used at 1km intervals to supplement

the information from the VRS dataset and calculate the approximate percentage of the segment with these barriers present.

- For minor junctions (T, X junctions and roundabouts), the counts produced by the TII base dataset were replaced with junction counts from the PRIME 2 data, as this were identified to be more accurate. When buffering (combining) the GIS layers, rounded ends were used to capture roundabouts and other junctions at the ends of segments.
- For major junctions, off ramp start points and on ramp end points in the PRIME 2 data were used to count the number of slip roads on each segment. If there was an on and an off slip within a segment (which is likely as carriageways are combined), this will count as two major junctions.

The full list of variables included in the modelling is documented in Table 8.

2.2.2 Method and results

In more detail, the segmentation used the following approach:

1. For each of the four road types, the base network was divided into 10m points, and each point assigned the nearest two-way AADT and radius of curvature.
2. Segments were identified using a combination of AADT, radius of curvature and changes in the number of lanes. Potential locations for segment division were identified by:
    a. identifying differences in AADT between adjacent points that were over a threshold.
    b. identifying points at which the number of lanes changed.
    c. radius of curvature lines were created by highlighting all points deemed to be curved (i.e. those over a certain threshold), merging the curved points into a line and then obtaining the ends of these lines.
3. The thresholds for each of the three parameters were set depending on the type of network, for example, Motorway network used an AADT difference of 5,000 while single carriageway networks used 1,000. See Table 2 for a full list of thresholds.

Table 2: Segmentation parameter Thresholds

Network Type	AADT	Radius of Curvature (km)	Lane Change	Max Length (km)
Motorway	5,000	3	1	5
Dual Carriageway	5,000	3	1	5
Single Carriageway	1,000	5	1	5
Legacy Road	1,000	5	1	5

4. A minimum segment length was applied and any identified segments less than 200m were combined with adjacent sections.
5. Any segments above the set maximum length threshold were divided into smaller sections of equal length (i.e. a segment of 7km was divided into two segments of 3.5km).
6. A validation process was used to fine tune the thresholds, ensuring that the distribution of collisions and segments lengths was appropriate for modelling (e.g. minimising the number of segments with zero collisions).
7. Once the segments were created, the other data sources were linked to these segments (using a buffer of 50m to capture all relevant data). The variables for the modelling (see Table 8) were then calculated. For example, this involved:
    a. Calculating a mean of the 10m AADT values.
    b. Counting the number of junctions in order to calculate a junction density.
    c. Identifying the minimum (tightest) radius on the segment.
    d. Combining continuous scale responses into categorical responses (e.g. median width).
8. Any segments with missing data for the variables of interest were removed prior to modelling.

2.2.2.1 Motorway network

Table 3 outlines the mean, median, minimum and maximum values for the 371 motorway segments. Of these segments, 25 (7%) had zero collisions.

Table 3: Properties of the motorway segmentation

Variable of interest	Mean	Median	Min	Max	Categorical variable counts
AADT	14,233	8,854	3,644	72,066	
Segment length (km)	2.6	2.1	0.2	5.0	
Collisions per segment	14.7	10	0	234	
Gradient	1.3	1.2	0.2	5.3	
Crossfall	1.6	1.5	1.0	3.1	
Radius	0.9	0.8	0.04	4.0	
CSC % (skid resistance)	97%	100%	26%	100%	
Urban/rural (categorical)	-	-	-	-	299 rural 12 urban 60 mixed
HGV %	11%	11%	3%	52%	

Variable of interest	Mean	Median	Min	Max	Categorical variable counts
Minor junction density (per km)	0.2	0	0	4.2	
Major junction density (per km)	1.3	0.4	0	15.9	
Hard shoulder width (categorical)	-	-	-	-	248 relaxation 128 wide
Median width (categorical)	-	-	-	-	11 narrow 187 standard 173 wide
M50 flag (categorical)	-	-	-	-	330 no 41 yes
Verge barrier %	25%	0%	0%	100%	

2.2.2.2 Dual carriageway network

Table 4 outlines the mean, median, minimum and maximum values for the 190 dual carriageway segments. Of these segments, 10 (5%) had zero collisions.

Table 4: Properties of the dual carriageway segmentation

Variable of interest	Mean	Median	Min	Max	Categorical variable counts
AADT	17,171	13,110	2,000	50,051	
Segment length (km)	1.4	0.9	0.2	5.0	
Collisions per segment	23.8	12.5	0	230	
Gradient	1.8	1.3	0.3	4.7	
Crossfall	1.7	1.6	1.0	3.2	
Radius	0.5	0.5	0	4.3	
CSC % (skid resistance)	82%	93%	5%	100%	
Urban/rural (categorical)	-	-	-	-	135 rural 50 urban 5 mixed
HGV %	7%	6%	1%	22%	
Minor junction density (per km)	2.5	0.6	0	37.1	
Major junction density (per km)	1.2	0	0	13.5	

Variable of interest	Mean	Median	Min	Max	Categorical variable counts
Verge barrier %	18%	0%	0%	100%	
Median barrier %	82%	100%	0%	100%	
Access Business density (per km)	1.3	0	0	102.5	
Access Commercial density (per km)	2.1	0	0	78.8	
Access Residential density (per km)	15.5	0.8	0	305.9	

2.2.2.3 Single carriageway network

Table 5 outlines the mean, median, minimum and maximum values for the 2,234 single carriageway segments. Of these segments, 329 (15%) had zero collisions.

Table 5: Properties of the single carriageway segmentation

Variable of interest	Mean	Median	Min	Max	Categorical variable counts
AADT	4,389	4,040	278	14,038	
Segment length (km)	0.9	0.6	0.2	5.0	
Collisions per segment	8.2	4.0	0	220	
Gradient	1.8	1.6	0.2	7.4	
Crossfall	1.6	1.6	0.3	3.7	
Radius	1.6	0.9	0.02	10	
CSC % (skid resistance)	53%	54%	0%	100%	
Urban/rural (categorical)	-	-	-	-	1,750 rural 409 urban 75 mixed
HGV %	9%	8%	1%	51%	
Minor junction density (per km)	2.9	2.0	0	28.0	
Major junction density (per km)	0.01	0	0	11.6	
Access Business density (per km)	2.3	0	0	360.7	
Access Commercial density (per km)	3.1	0	0	281.7	
Access Residential density (per km)	12.0	3.4	0	289.0	
2+1 or 1+1 flag (categorical)					55 Yes

2.2.2.4 *Legacy road network*

Table 6 outlines the mean, median, minimum and maximum values for the 1,703 legacy segments. Of these segments, 319 (19%) had zero collisions.

Table 6: Properties of the legacy network segmentation

Variable of interest	Mean	Median	Min	Max	Categorical variable counts
AADT	1,965	1,886	195	6,702	
Segment length (km)	1.2	0.7	0.2	5.0	
Collisions per segment	5.5	3.0	0	116	
Gradient	2.6	2.4	0.2	8.8	
Crossfall	1.6	1.6	0.4	4.1	
Radius	1.2	0.6	0.04	10	
CSC % (skid resistance)	71%	84%	0%	100%	
Urban/rural (categorical)	-	-	-	-	1,453 rural 233 urban 17 mixed
HGV %	9%	8%	0%	31%	
Minor junction density (per km)	2.7	2.0	0	20.6	
Major junction density (per km)	0	0	0	0	
Access Business density (per km)	2.7	0.3	0	208.6	
Access Commercial density (per km)	2.9	0	0	268.7	
Access Residential density (per km)	10.6	4.0	0	314.1	

2.3 APM development

The modelling process described below is based on the methodology proposed in the interim report (Chowdhury, et al., 2022).

Since the number of collisions on some sections is zero, both Generalised Linear Models (GLMs) and zero-inflated models were tested on the data. In all cases, the zero-inflated models were found to fit the data substantially better, resulting in improved prediction estimates. Zero-inflated models consist of a two-stage modelling process:

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 Negative Binomial or Poisson model to model the non-zero count data. Within this model, "sampling zeros" are modelled for those observations which are exposed to the risk but do not report experience of the outcome during the study.

In the context of collision data, the first modelling process can be thought of as modelling the segments with such low collision risk (e.g. with low exposure or few risky factors being present) that you will only get zero collision counts on these. The second models the segments with non-zero collision risk. These are sections where collisions are likely to occur, however some of these segments will not have a collision(s) reported during the sampling period. Note that in these models, an additional coefficient called theta (θ) is also calculated alongside the coefficients for the variables of interest.

For each road type, the first step is to determine the simple base model. This was developed using two of the segmentation variables. These were AADT and segment length. These variables are known from the literature to be the parameters that (almost always) account for the greatest variability in collision occurrence in these GLMs (Table 7). The number of lanes was not included as a variable in the base model as it was highly correlated with AADT. This meant that when it was tested for inclusion in this model, number of lanes was not a significant predictor of collisions.

Table 7: Variables included in the base model (all models)

Models	Variable	Description
ALL	AADT	Average AADT
ALL	Segment length	Length of segment in kilometres

Previous models developed by others (see Chowdhury et. al, (2022) for details) have included these variables in the model as either the power or exponential form. These models were compared to each other using their Akaike Information Criterion (AIC)[1] scores to determine which form generated a better fit for these data. The models were also used to predict the number of collisions if the value of the variable was zero. This served as a logic check since a segment with a flow of zero should be predicted to have zero collisions. The form of the variable was chosen based on the results of the AIC score comparison and the logic check; results for each of the different road types are summarised in Sections 2.3.1 to 2.3.4 below.

Both Poisson and Negative Binomial zero-inflated base models were created. These models were compared to each other using a likelihood ratio test[2] to determine which type was better suited for the data. For all four road types, negative binomial models were determined to be the best fit.

[1] The AIC is a statistical method used to assess the goodness of fit of a model. It allows comparison of models to determine which one best explains the data.

[2] The likelihood ratio test assesses the goodness of fit of two competing models.

For zero-inflated models, different predictor variables can be used for each of the two stages in the modelling. 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. Segment length was also considered for inclusion here, to align to the base model variables, but due to the way the segmentation exercise was completed, many of the short network segments contain junctions, where we would expect the collision risk to be higher.

In the text that follows, the process to decide which variables should be used in the final model relate to the second stage of the modelling (the negative binomial model) – this is the main output of interest (and the coefficients for these variables are presented in Sections 2.3.1 to 2.3.4). These final models were developed from the base models by testing a number of variables for inclusion in the model and deciding which of these best explain the collision counts. All the possible variables that were assessed are shown in Table 8. There were some variables listed in the Interim report (Chowdhury, et al., 2022) which were ruled out of the modelling process following further investigation; these are detailed in Appendix A.

Table 8: Variables tested for inclusion in the models

Models	Variable	Description
All	Gradient	Maximum absolute gradient over segment
All	Crossfall	Mean crossfall over segment
All	Radius	Minimum radius on segment
All	CSC %	Percentage of the segment where Characteristic SCRIM Coefficient (CSC) values are over threshold. The SCRIM coefficient measures the skid resistance of the network; minimum thresholds are defined for each road type.
All	Urban/Rural	Classification into Rural, Urban, and Mixed
All	HGV %	Percentage of vehicles that are Heavy Goods Vehicles (HGVs)
All	Minor junctions	Minor junction density (number of minor junctions[3] per kilometre)

[3] This includes T-junctions, crossroads, and roundabouts.

Models	Variable	Description
Motorway, dual, single[4]	Major junctions	Major junction density (number of major junctions[5] per kilometre)
Motorway, dual, single[4]	Junctions	Overall junction density (minor + major junctions per kilometre)
Motorway	Hard shoulder width	For each carriageway the hard shoulder width is classified into: Narrow (<1m), Relaxation (1m-2.49m), Wide (≥2.5m). The minimum of these categories was selected for each segment.
Motorway	Median width	The median width is categorised into: Narrow (<2.49m), Standard (2.5m-4.79m), Wide (≥4.8m)
Motorway	M50 flag	Flags segments which are on the M50 (this was identified as a potential outlier by TII)
Motorway, dual	Verge barrier %	Percentage of the segment with a verge barrier on the nearside. Note: as segments contain both carriageways, if only one side of the carriageway has a verge barrier, this would be recorded as 50%.
Dual	Median barrier %	Percentage of the segment with a median barrier recorded. Note: this variable was not used for the motorway model as by definition, all motorway segments will have a median barrier.
Dual, single, legacy	Access Business	Access density to business premises (number of business accesses per kilometre)
Dual, single, legacy	Access Commercial	Access density to commercial premises (number of commercial accesses per kilometre)
Dual, single, legacy	Access Residential	Access density to residential premises (number of residential accesses per kilometre)
Single	2+1 or 1+1 flag	Flags segments which are one of these two road types

The variables that were added to each model were selected through a stepwise variable selection process, according to how significant they were (those with the best p-values were added first). Change in Akaike's Information Criterion (AIC) and Bayesian Information

[4] Note, there were no major junctions recorded on legacy roads so this variable was excluded from the modelling.

[5] This includes major roundabouts and slip roads.

Criterion (BIC)[6] values were also taken into account. At each step of the process, the following method was applied:

1. *n* models were created, where *n* was the number of variables left under consideration. Each model added one variable to those selected during the previous iteration.
2. The minimum p-value of these new variables in their models was determined, if the minimum value was greater 0.05 (non-significant) the process terminated.
3. The AIC and BIC values of the models with this new variable were compared to those of the model without this new variable. The variable was added to the model only if both AIC and BIC values decreased.
4. The process then updated list of selected variables and variables under consideration and looped back to step 1.

Throughout the process, variables with strong correlations between them were not included together as this would lead to multicollinearity[7] issues.

The output from the final model is a list of coefficients which describes how each variable influences collision risk (Table 9 summarises these relationships).

Table 9: Interpretation of coefficient values on collision risk

Coefficient (b)	
b > 1	For increasing values of the variable, the number of collisions will increase, at an increasing rate
b = 1	For increasing values of the variable, the number of collisions will increase, at a constant (or linear) rate
0 < b < 1	For increasing values of the variable, the number of collisions will increase, at a decreasing rate

[6] Both AIC and BIC are used to assess the goodness of fit of a model and compare between models to see which one best explains the data. The BIC also penalises models based on the number of parameters in the model, favouring 'simpler' models with fewer variables to ensure the model is not overfitted to the data. Overfitting can mean the model is only relevant to the data set it was built using, and irrelevant to any other data sets; the intention is to use the model outputs in the tool, so it is important this is not the case. As a result, a decision was made to use both the AIC and BIC in model selection, even though this may result in a more conservative model.

[7] Multicollinearity is an issue in statistically models where multiple explanatory variables are highly correlated to each other. This results in less reliable statistical inferences as the modelling will be unable to assign variance clearly to specific variables. All pairs of variables included in the model were therefore tested for correlation and for those where this was identified as 'strong' (Pearson's correlation coefficient >0.5), only one of the variables was added during the modelling (assuming it met the criteria of significance and reduction in AIC/BIC).

Coefficient (b)	
b = 0	There will be no change in the number of collisions with increasing values of the variable
b < 0	For increasing values of the variable, the number of collisions will decrease

For each of the final models selected, a comparison to the base model on the following measures is presented: AIC, BIC and McFadden's R-Squared. For AIC and BIC, a smaller number represents a better model. For McFadden's R-squared value, higher numbers are better, and values between 0.2 and 0.4 represents a very good model fit (Hensher & Stopher, 1979).

The final step of the modelling process was to assess the predictive performance of the model. This is achieved by using K-fold cross-validation. A K value of 10 was used as this is standard practise. In K-fold cross validation the data is split into K subsets of equal size. This allows for 10 combinations of training and tests sets such that 9 subsets are combined to form a training set and one subset is used as the test set. The model is trained on the training set, then the predictive performance of the model is measured using the test set and some selected metrics. The metrics used were the Mean Absolute Deviance (MAD)[8] and the Mean Squared Prediction Error (MSPE)[9]. The average values for these metrics across the ten test sets provide an estimate of the predictive performance of the model.

During the process of developing the models, some improvements to the data were made. These are documented in Appendix B.

The following sections present the results of this modelling exercise for the main response variable of interest: all collisions (including material damage only). Additional modelling for collisions involved casualties who were killed or seriously injured (KSI) only are included in Appendix C.

2.3.1 Motorway model

2.3.1.1 Base model

The base model included AADT and segment length variables both in power form.

[8] MAD is estimated by subtracting the actual collision values from the predicted values, converting it to an absolute error and calculating the average. While this metric is the easiest to explain, the main drawback of this measure is that it averages out the error across the entire dataset which does not necessarily present a true picture of the prediction error across the full range of values. As a result, MAD is less accurate for outliers but better for 'normal' observations.

[9] The main advantage of the MPSE is that it is more sensitive to large outliers compared to MAD; however, it might be less accurate for 'normal' observations.

The likelihood ratio test result indicated that the zero-inflated negative binomial model had significantly better goodness-of-fit than the equivalent Poisson model with $p<0.01$.

2.3.1.2 Full model

Over the course of the stepwise variable selection process the variables shown in Table 10 were selected for inclusion based on their p-values, AIC and BIC values.

Table 10: Variables included in motorway model for all collisions (variables of interest in black, others also included in the model in grey)

Model stage	Variable	Coefficient	p-value
Count model	*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
	Log(theta)	*1.647*	*p<0.001*
Zero-inflation model	*Intercept*	*44.478*	*0.008*
	Log(AADT)	-5.468	0.005

The direction of the coefficients (positive or negative) is as expected:

- As the segment length or AADT increase, the positive coefficients indicate that collision risk increases.
    - The segment length coefficient is less than 1 which indicates that if the segment length is doubled, the collision risk is increased, but not as much as doubled.
    - The AADT coefficient is slightly greater than 1, suggesting that as the flow increases, the collision risk increases at an increasing rate. This result has been found in previous studies for motorways (e.g. the M25 Controlled Motorways safety benefit report for the GB government found an exponent of 1.663).
- Increases in gradient, and increases in the % of vehicles which are HGVs, both increase collision risk, possibly due to increased speed differentials and braking performance between vehicle types.
- As the radius increases (i.e. the road becomes less bendy), the collision risk decreases.

The goodness of fit measures for this final model are compared to the base mode in Table 11. AIC and BIC have fallen, and the R-squared value has increased, suggesting that the final model is better than the base model for predicting collision risk. However, the final McFadden's R-squared value is less than 0.2, suggesting this model could be improved.

Table 11: Goodness of fit measures for the motorway model for all collisions

	Base model	Final model
AIC	2,343	2,301
BIC	2,367	2,336
McFadden R-Squared	0.16	0.17

2.3.1.3 Model prediction accuracy

The model predictive performance was assessed using 10-fold cross validation. The model prediction accuracy is assessed through a comparison of the actual collision numbers to those predicted by the model (Figure 4Figure) and two evaluation metrics: the Mean Absolute Deviance (MAD) and Mean Squared Prediction Error (MSPE). These values are calculated for all ten of the samples and the mean value is presented in Table 12; lower error values indicate better model fit.

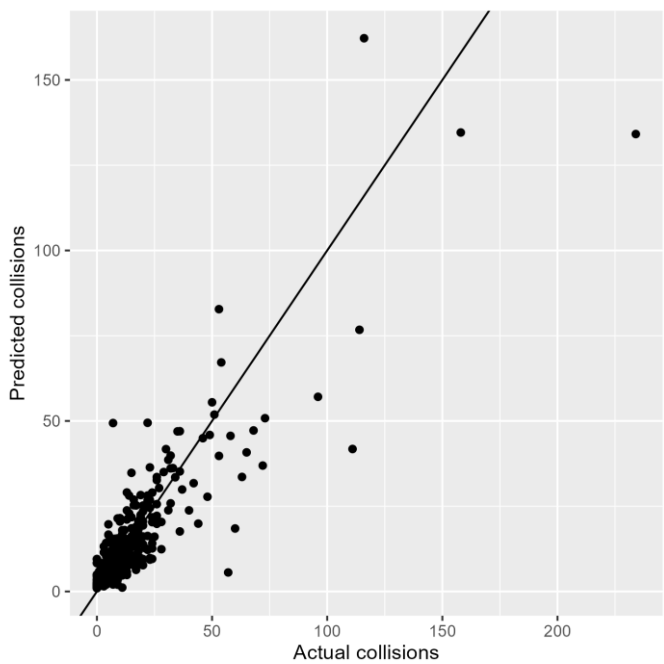


Figure 4: Assessment of model predictions against actual collision numbers for the motorway all collisions model

Table 12: Prediction accuracy for the motorway model for all collisions

	Final model
Mean (MAD)	5.7
Sqrt(mean (MSPE))	10.8

The results indicate that although there is some under and over predictions (results lie below and above the line), the predictions are relatively close to the line. On average, the difference in collisions between the actual and predicted for each segment is around 5 to 11 collisions.

2.3.2 Dual carriageway model

2.3.2.1 Base model

The base model included AADT and segment length variables, both in power form.

The likelihood ratio test result indicated that the zero inflated negative binomial model had significantly better goodness-of-fit than the equivalent Poisson model with p<0.01.

2.3.2.2 Full model

Over the course of the stepwise variable selection process the variables shown in Table 13 were selected for inclusion based on their p-values, AIC and BIC values.

Table 13: Variables included in dual carriageway model for all collisions (variables of interest in black, others also included in the model in grey)

Model stage	Variable	Coefficient	p-value
Count model	*Intercept*	*-6.872*	*p<0.001*
	Log(segment length)	0.597	p<0.001
	Log(AADT)	1.144	p<0.001
	Median Barrier %	-1.020	p<0.001
	Radius	-0.697	p<0.001
	Major Junctions	-0.118	p<0.001
	Access Commercial	0.019	0.003
	Log(theta)	*0.838*	*p<0.001*
Zero inflated model	*Intercept*	*6.118*	*N/A*
	Log(AADT)	-1.749	0.79 (ns)

Except for major junctions, the direction of the coefficients (positive or negative) is as expected:

- As the segment length or AADT increase, the positive coefficients indicate that collision risk increases.
  - The segment length coefficient is around 0.6 so as the length is doubled, the increase in collisions is less than double.
  - The magnitude of the AADT variable is similar to that seen for motorways (Table 10) and comparable to other studies.
- A greater proportion of median barriers decreases the risk on a segment. Whilst presence of a barrier may not in fact alter the number of collisions, barriers are known to reduce the collision severity, primarily by reducing head on collisions.
- As the radius increases (i.e. the road becomes less bendy), the collision risk decreases.
- As the density of major junctions (slip roads) increases, collision risk decreases. This result is counterintuitive since junctions are known to increase collisions. However, in the absence of a speed variable in the model (see Appendix A for the reasons for this), this may be acting as a proxy for the lower speeds typically observed around junctions.
- Unlike for junctions, as the density of commercial access points increase, collision risk increases which aligns with expectations for this variable.

The goodness of fit measures for this final model are compared to the base mode in Table 14. AIC and BIC have fallen, and the R-squared value has increased, suggesting that the final model is better than the base model for predicting collision risk. However, the final McFadden's R-squared value is quite a bit less than 0.2, suggesting this model is not particularly strong.

Table 14: Goodness of fit measures for the dual carriageway model for all collisions

	Base model	Final model
AIC	1,520	1,398
BIC	1,539	1,430
McFadden R-Squared	0.05	0.13

2.3.2.3 Model prediction accuracy

The model predictive performance was assessed through a comparison of the actual collision numbers to those predicted by the model (Figure 5Figure) and two evaluation metrics: MAD and MSPE (Table 15).

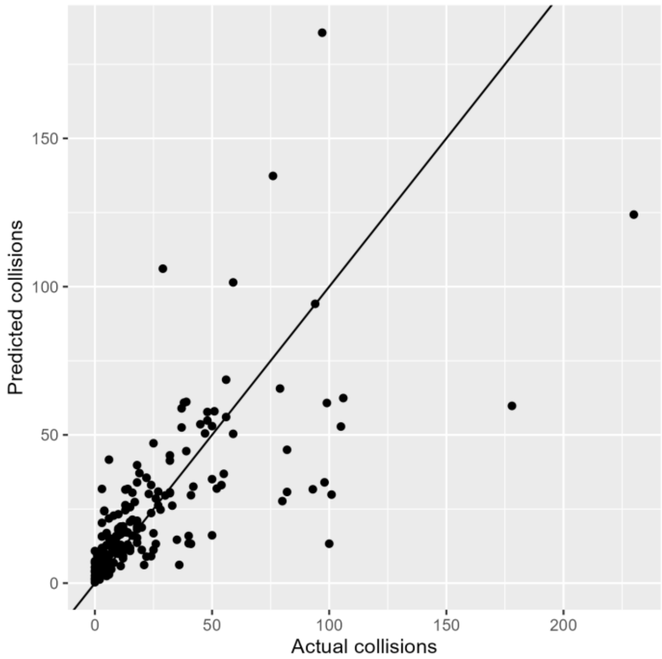


Figure 5: Assessment of model predictions against actual collision numbers for the dual carriageway all collisions model

Table 15: Prediction accuracy for the dual carriageway model for all collisions

	Final model
Mean (MAD)	12.7
Sqrt(mean (MSPE))	22.7

The results indicate that there is some under and over prediction (results lie below and above the line). On average, the difference in collisions between the actual and predicted for each segment is around 13 to 23 collisions, which is higher than for the motorways model and aligns to the lower goodness of fit value for this model.

2.3.3 Single carriageway model

2.3.3.1 Base model

The base model included AADT and segment length variables, both in power form.

The likelihood ratio test result indicated that the zero-inflated negative binomial model had significantly better goodness-of-fit than the equivalent Poisson model with $p<0.01$.

2.3.3.2 Full model

Over the course of the stepwise variable selection process the variables shown in Table 16 were selected for inclusion based on their p-values, AIC and BIC values.

Table 16: Variables included in single carriageway model for all collisions (variables of interest in black, others also included in the model in grey)

Model stage	Variable	Coefficient	p-value
Count model	*Intercept*	*-5.888*	*p<0.001*
	Log(segment length)	0.841	p<0.001
	Log(AADT)	0.877	p<0.001
	Minor Junctions	0.132	p<0.001
	Access Commercial	0.015	p<0.001
	Gradient	0.169	p<0.001
	Radius	-0.073	p<0.001
	CSC %	-0.186	p<0.001
	Log(theta)	*0.777*	*p<0.001*
Zero inflated model	*Intercept*	*1.782*	*0.983 (ns)*
	Log(AADT)	-1.496	0.885 (ns)

The direction of the coefficients (positive or negative) is as expected:

- As the segment length or AADT increase, the positive coefficients indicate that collision risk increases.
  - The segment length coefficient is similar in magnitude to the motorway model (Table 10). This coefficient is reasonably close to 1 which indicates that as the segment length is doubled, the collision risk is almost doubled collision risk.
  - Unlike for the motorway and dual carriageway models, the magnitude of the AADT coefficient is less than 1 for single carriageways. Values below one are not uncommon in these studies.
- As the density of the minor junctions (T-junctions, X junctions and roundabouts) increases, the collision risk increases. The same is true for access to commercial premises, although the magnitude of this effect is smaller than for minor junctions.
- The characteristics of the road segments influence collision risk:
  - As the gradient increases, the collision risk increases (with a similar coefficient magnitude to the motorway model - Table 10).
  - As the radius increases (i.e. the road becomes less bendy), the collision risk decreases. The magnitude of this coefficient is smaller than for motorways and dual carriageways but, as outlined in Section 2.2.1, the calculations for

this variable are slightly different for single and legacy roads so cannot be directly compared.

- As the percentage of the segment where the CSC (skid resistance) values are over the threshold increases (i.e. more of the road meets the minimum criteria for skid resistance on this road type), collision risk decreases.

The goodness of fit measures for this final model are compared to the base mode in Table 17. AIC and BIC have fallen, and the R-squared value has increased, suggesting that the final model is better than the base model for predicting collision risk. However, the final McFadden's R-squared value is less than 0.2, suggesting this model could be improved.

Table 17: Goodness of fit measures for the single carriageway model for all collisions

	Base model	Final model
AIC	12,932	11,720
BIC	12,966	11,783
McFadden R-Squared	0.07	0.16

2.3.3.3 Model prediction accuracy

The model predictive performance was assessed through a comparison of the actual collision numbers to those predicted by the model (Figure 6Figure and Figure 7 with the three outliers with very large predictions removed) and two evaluation metrics: MAD and MSPE (Table 18).

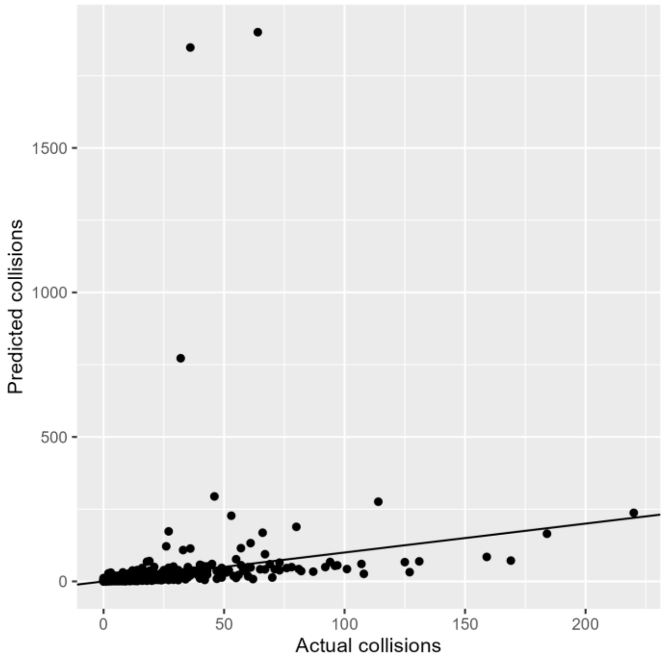


Figure 6: Assessment of model predictions against actual collision numbers for the single carriageway all collisions model

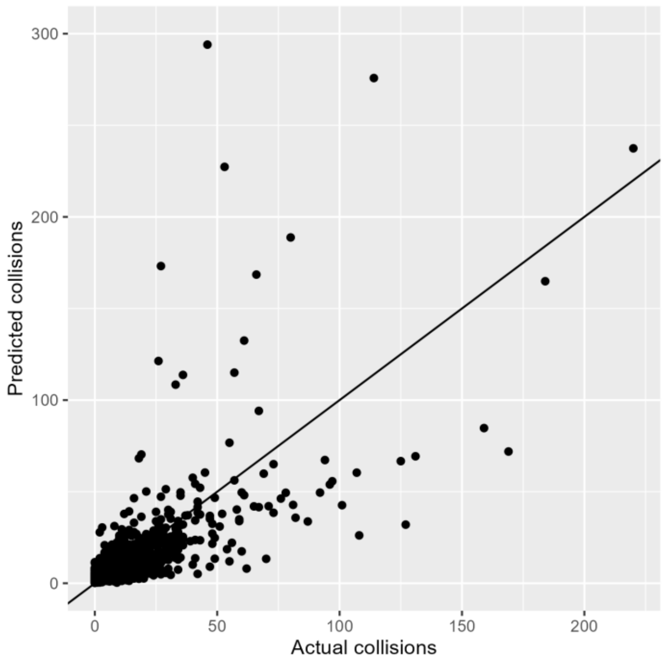


Figure 7: Assessment of model predictions against actual collision numbers for the single carriageway all collisions model (outliers removed)

Table 18: Prediction accuracy for the single carriageway model for all collisions

	Final model
Mean (MAD)	6.7
Sqrt(mean (MSPE))	58.0

The results indicate that there are three significant outlying predictions (shown in Figure 6). Further investigation of these segments (Figure 8) identifies all three are located in the middle of a small town (Longford) on the high street with shops, restaurants and other commercial buildings being present. There are multiple surrounding car parks plus a college and a cathedral less than 200m from the centre. The area is also a junction for many trunk roads; the N5, N63 and N4 join in the centre of the town. As a result, the features of these sections which make the model predictions high are:

- High minor junction density
- High commercial access density
- Short segment length
- Low minimum radius (bendy sections).

Figure 8: Single carriageway model segments with very high collision predictions[Map sourced from Google (2023a)]

If these three results are excluded (Figure 7), the fit of the model appears to be much better, although, as with the other models, there is some under and over prediction (results lie below and above the line).

On average, the mean absolute deviance (MAD) is around 6 collisions, which is comparable to the figure for the motorways model and lower (better) than for dual carriageways. As outlined earlier, the MPSE figure is more sensitive to outliers than the MAD and hence this result is substantially bigger (worse) than for the motorway and dual models.

2.3.4 Legacy road model

2.3.4.1 Base model

The base model included AADT and segment length variables, both in power form.

The likelihood ratio test result indicated that the zero-inflated negative binomial model had significantly better goodness-of-fit than the equivalent Poisson model with p<0.01.

2.3.4.2 Full model

Over the course of the stepwise variable selection process the variables shown in Table 19 were selected for inclusion based on their p-values, AIC and BIC values.

Table 19: Variables included in legacy roads model for all collisions (variables of interest in black, others also included in the model in grey)

	Variable	Coefficient	p-value
Count model	*Intercept*	*-3.767*	*p<0.001*
	Log(segment length)	0.970	p<0.001
	Log(AADT)	0.680	p<0.001
	Minor Junctions	0.081	p<0.001
	Access Commercial	0.020	p<0.001
	CSC %	-0.298	p<0.001
	Radius	-0.052	p<0.001
	Gradient	0.054	0.003
	Log(theta)	*1.043*	*p<0.001*
Zero inflated model	*Intercept*	*-5.738*	*1.000 (ns)*
	Log(AADT)	-2.148	0.999 (ns)

The direction of these coefficients (positive or negative) is as expected:

- As the segment length or AADT increase, the positive coefficients indicate that collision risk increases.
    - The segment length coefficient is very close to 1 which indicates that as the segment length is doubled, the collision risk is almost doubled.
    - As with single carriageways (Table 16), the coefficient for AADT is below 1, meaning a doubling of flow will not result in a doubling of collision risk.

- As the density of minor junctions (T-junctions, X junctions and roundabouts) and commercial accesses increase, the collision risk increases. The magnitude of the commercial access effect is similar to single carriageways (Table 16) but the minor junction effect is smaller.
- As the percentage of the segment where the CSC (skid resistance) values are over the threshold increases (i.e. more of the road meets the minimum criteria for skid resistance on this road type), collision risk decreases. This is a larger effect than that observed for single carriageways (Table 16).
- As with all other road types, the characteristics of the road segments influence collision risk:
    - As the gradient increases, the collision risk increases (although the coefficient is smaller than for motorways (Table 10) and single carriageways (Table 16).
    - As the radius increases (i.e. the road becomes less bendy), the collision risk decreases. The magnitude of this coefficient is very similar to single carriageways (Table 16).

The goodness of fit measures for this final model are compared to the base mode in Table 20. AIC and BIC have fallen, and the R-squared value has increased, suggesting that the final model is better than the base model for predicting collision risk. However, the final McFadden's R-squared value is less than 0.2, suggesting this model could be improved.

Table 20: Goodness of fit measures for the legacy roads model for all collisions

	Base model	Final model
AIC	8,404	7,811
BIC	8,436	7,871
McFadden R-Squared	0.11	0.17

2.3.4.3 Model prediction accuracy

The model predictive performance was assessed through a comparison of the actual collision numbers to those predicted by the model (Figure 9 and Figure 10 with the two outliers with very large predictions removed) and two evaluation metrics: MAD and MSPE (Table 21).

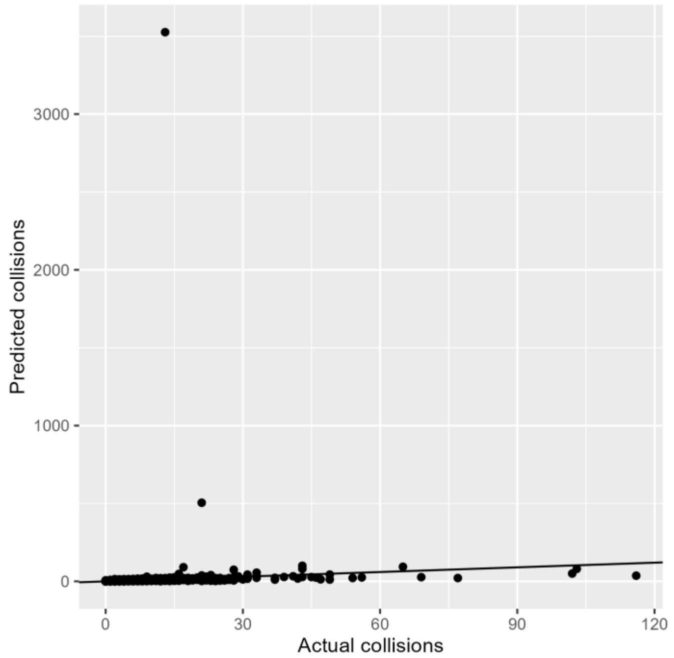


Figure 9: Assessment of model predictions against actual collision numbers for the legacy roads all collisions model

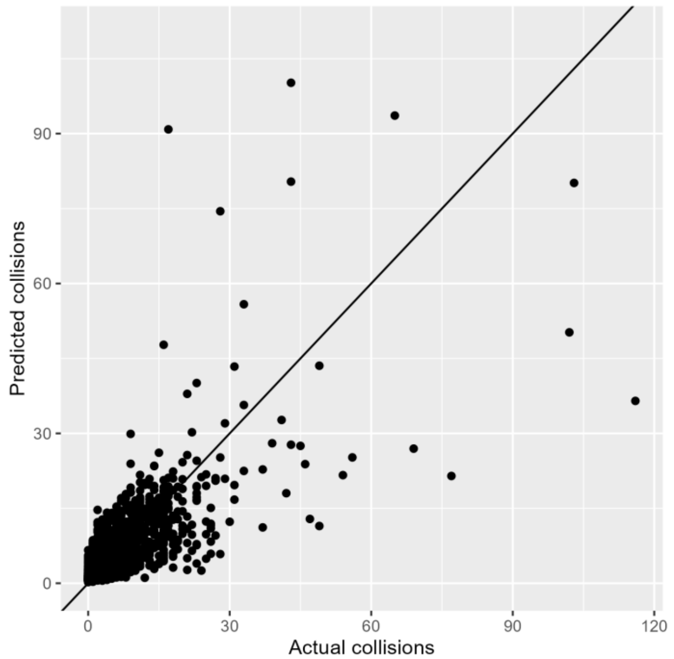


Figure 10: Assessment of model predictions against actual collision numbers for the legacy carriageway all collisions model (outliers removed)

Table 21: Prediction accuracy for the legacy roads model for all collisions

	Final model
Mean (MAD)	5.3
Sqrt(mean (MSPE))	86.0

The results indicate that there are two substantial outlying predictions (shown in Figure 9Figure). Further investigation of these segments (Figure 11) shows for the segment with the prediction of over 3,500 collisions (near the town of Boyle); the following characteristics make this segment high risk:

- High gradient
- Low minimum radius (bendy section)
- Minimum CSC % (0%, because all the values on this segment are below the threshold)
- High minor junction density
- High commercial access density (located in the centre of a village with many surrounding businesses and shops)
- Short segment length

For the other segment with 500+ collisions predicted (near the town of Listowel), the last three bullet points above also apply.

Segment with over 3,500 predicted collisions

Segment with over 500 predicted collisions

Figure 11: Legacy carriageway model segment with very high collision predictions [Maps source from Google (2023b) and (2023c)]

If these two results are excluded (Figure 10), the fit of the model appears to be much better, although, as with the other models, there is some under and over prediction (results lie below and above the line).

On average, the mean absolute deviance (MAD) is around 5 collisions, which is comparable to the figure for the motorways and single carriageway models and lower (better) than for dual carriageways. As outlined earlier, the MPSE figure is more sensitive to outliers than the MAD and hence this result is substantially bigger (worse) than for the other models.

2.4 Model limitations and future improvements to the models

As outlined in Appendix B, several data improvements were made when developing these models. However, there are also a number of acknowledged limitations of the modelling and subsequent improvements which should be considered for future model improvements:

6. One of the largest limitations of the models developed is that some key variables are known to be missing from the variables considered for inclusion in Table 8. One example of this is vehicle speed (Appendix A details why this could not be included).

   The absence of variables which are known to influence collision risk means the model may be 'underspecified'[10], which can lead to biased coefficients and predictions. As a result, the coefficients presented in this report should be treated as indicative and the models should be updated as new data sources are available.

7. It is recommended that TII consider how suitable speed data might be collected; as a minimum this should include mean and 85th percentile speeds to enable the range of operating speeds to be understood.

8. Additional variables which could be considered for future modelling if suitable data could be collected include:

   a. Road user flow data by road user type (e.g. motorcycles, pedestrians, pedal cycles) – HGV percentage was included as a variable to measure the prevalence of these vehicle types but understanding the presence of vulnerable road users would be a useful addition to the explanation of collision risk, in particular for KSIs.

   b. Presence of roadside and carriageway features including rumble strips, street lighting, pedestrian crossing facilities and cycle facilities.

9. 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.

10. Since the conditions on each carriageway can differ within a segment the modelling presented here is limited because carriageways had to be combined. If the collision

[10] This can result in a model with coefficients which are unstable (i.e. change substantially as additional variables are added to the model) and wider than necessary confidence intervals around these coefficients.

data could be reliably assigned to a carriageway, this would enable separate segmentations per carriageway and more reliable models to be developed.

11. Accurate assignment of collisions to specific roads would also enable better assignment of collisions and junctions to modelling segments, eliminating some of the challenges with assignment where segments cross over (e.g. where one road crosses over the top of another). At present, there are some locations where collisions may be incorrectly assigned to the segment because it is not possible to determine which road the collision occurred on.

3 Task 5 – stakeholder engagement and tool planning

This section presents a summary of the results of two tasks led by Arup:

1. An online survey with Road Safety Engineers that was prepared and utilised to gather opinions and views on a Transport Infrastructure Ireland (TII) Collision Reduction Calculator (Section 3.1).
2. Workshops held with Local and Regional Engineers which facilitated a more detailed discussion on what is needed from the tool and how this might be used by end users (Section 3.2).

The implications for the design of the tool are summarised in Section 3.3.

3.1 Online survey results

An online survey was prepared by Arup and TRL and circulated to those taking part in the workshops. The objective of the on-line survey was to gain an understanding of the following:

- Current processes for assessing schemes,
- How Collision Modification Factors (CMFs) are being utilised,
- The typical countermeasures used,
- How is the effectiveness of a proposed scheme used.

There were 14 responses received to the on-line survey, full results of the survey are presented in Appendix D. The results showed that:

- There is a varied approach to assessing road safety improvement schemes including use of the collision statistics, local knowledge, site visits, Google Street View, gathering of Road Safety Audit (RSA) data (if it can be obtained – GDPR is an issue), traffic data (from TII sites or temporary counts), drones/photos/camera surveys, information from local Gardai and/or residents.
- 11 out of 14 respondents said they did use CMFs to assess road safety improvements. The majority of these were sourced from TII and CMF Clearinghouse.
- The most common countermeasures were reported to be:
  - Signing and Lining (12 responses)
  - Improved visibility (10 responses)
  - Active Travel Measures (10 responses)
  - Traffic Calming (9 responses)
- The effectiveness (safety benefits) of a scheme is assessed through calculations of the expected reduction in collisions, First Year Rate of Return (FYRR) calculations, before and after analysis (collisions or speed) or changes in local perceptions.

3.2 Workshops with TII

A series of workshops were held with TII, Regional Road Safety Engineers (RRSE) from TII and National Roads Design Offices (NRDOs).

Building on the on-line survey, the workshops were intended to:

- Gather input from those likely to use the tool, the format, inputs, outputs, etc. to develop a useful tool with a stronger end user experience.
- Allow the contributors to provide suggestions for additional features and how the tool would work.

There were nine attendees across the two workshops. The key findings from the workshops can be summarised as follows:

1. **Combining countermeasures is important.** When combining CMFs, simplicity is important as the output from the tool needs to be easy for the user to understand. Upper and lower bounds on the CMFs could be useful to understand the range of expected effectiveness of the interventions, but it should be clear which figure to use for appraisal purposes.

   Most schemes have multiple interventions, so it is important that the tool enables this to be captured by the inputs.
2. **First Year Rate of Return (FYRR) is a useful metric for TII.** If this calculation can easily be added to the tool as an output, this would be of great benefit to the end users of the tool.
3. **The audience for the tool have a technical background, although road safety may not their field of expertise.** The tool should include notes, clear signposting and drop-down menus are required.
4. **CMFs should be referred to as a "Collision Reduction %".** The term 'CMF' is not generally well liked or understood by the users of the tool; a 'collision reduction %' is the most useful output. Understanding how the countermeasure leads to a reduction in the severity of collisions (e.g. fewer KSIs and more damage only) in the tool would also be useful.
5. **A feedback loop with engineers is needed during tool development.** Those involved in the workshops expressed a view that they would like to be able to input during development of the tool, particularly on which countermeasures are included.
6. **A variety of countermeasures that were suggested for inclusion** including VRU crossings, roundabouts, urban landscaping, lane narrowing, traffic calming (channelised islands), school zones - lining and signing, rural bus stop improvements and junction visibility improvements.
7. **The tool would be useful more widely on the Irish road network, not just on TII managed roads.** The implications of this should be considered during tool development and any caveats on its use in this situation made clear in the tool. If the tool is not suitable for use on other road types, this needs to be stated in the tool

guidance. Users are likely to be regional inspection engineers, area engineers and others.

It would be useful to consider if an option to input your own CMF or add countermeasures later could be included to future proof the tool.

8. **Previous work carried out in 2020 on CMF research and guidance for application should be reviewed.** This project should build on this guidance as this work reviewed Clearinghouse CMFs[11] and their application to Ireland.
9. **A suggestion was made to utilise only the top (4 and 5 star) CMFs from Clearinghouse and average these for use in the tool.**

3.3 Tool design

From this engagement, several key decisions were made:

- It was confirmed that the tool would be developed in Excel with relatively simple functionality for the user (and no macros).
- The output would include the FYRR and be easy to save and store (via a print out) for future reference to support the business case for intervention.
- The tool would be available for download on the TII website and would include guidance on its use within the tool itself, rather than requiring separate documentation.
- On the main calculations page, the term 'collision reduction percentage' would be used in the tool, rather than 'CMF'. The term 'CMF' will be used on a separate tab with more detailed calculations only.
- A central (average) estimate for the collision reductions would be presented on the main tab. Supporting pessimistic and optimistic bound would be presented on the more detailed calculations tab for users who want this level of detail.
- Following the workshop, the suggestion to only use 4- and 5-star CMFs from Clearinghouse was discussed further with TII and it was decided this would severely limit the capabilities of the tool. A decision was made to use 3-, 4- and 5-star CMFs.
- A decision was made to refer to the tool as the 'Collision Reduction Calculator' so users are clear on what the tool can be used to do.

[11] Clearinghouse (https://www.cmfclearinghouse.org/) is a repository for CMFs from internationally published studies. It provides a searchable database of CMFs along with guidance and resources on using CMFs in road safety practice. Other databases (e.g. the PRACT repository) also exist but Clearinghouse is the most up to date source for these data.

4 Task 6 – Collision Reduction Calculator development

This section gives a high-level overview of the Collision Reduction Calculator functionality (Section 4.1), summarises the Irish countermeasures and CMFs included in the calculator from the modelling (Section 4.2) and presents some suggestions for future improvements to the tool (Section 4.3).

The calculator is available to download from the TII Publications website: https://www.tiipublications.ie/.

4.1 Overview of the Collision Reduction Calculator

Following the survey and workshop a full scope and process flow was developed and agreed with TII. Figure 12 presents a high-level summary of the Collision Reduction Calculator process flow. Full details of the user process flow and examples of the calculations are presented in Appendix E.

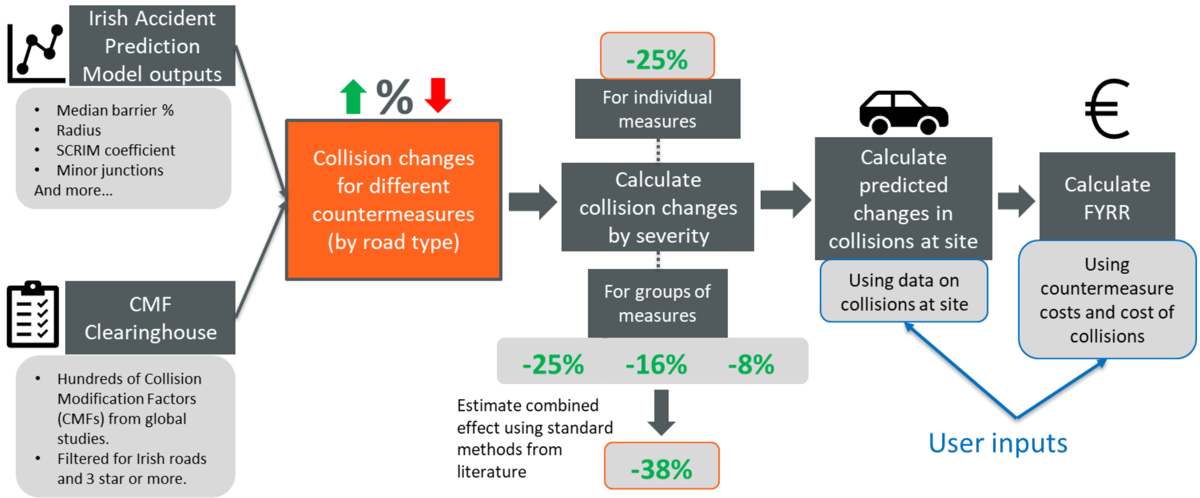


Figure 12: Summary of the process flow for the Collision Reduction Calculator

The inputs to the calculator are the CMFs from the Irish APMs (outlined in Section 4.2) and CMFs from the Clearinghouse website. Filters were applied to the CMFs from Clearinghouse to utilise only the highest quality studies (see Appendix E.1 for more details).

The user selects the road type of interest: motorway, dual carriageway, single carriageway or legacy road, and this further filters the countermeasures available for selection to only those which are applicable to the road type of interest (the definitions of each road type used in Clearinghouse are summarised in Appendix E.2).

The user selects countermeasures of their choice from the list, or inputs their own, and the associated collision changes by severity are presented and combined (where necessary, using standard methods from the literature[12]) to calculate predicted changes in the collision numbers.

The user inputs information on the historical collisions at the site and the estimated cost of these countermeasures, and the First Year Rate of Return (FYRR) is presented as a percentage as the key output. A screenshot of the results section of the main calculator page of the tool is given in Figure 13.

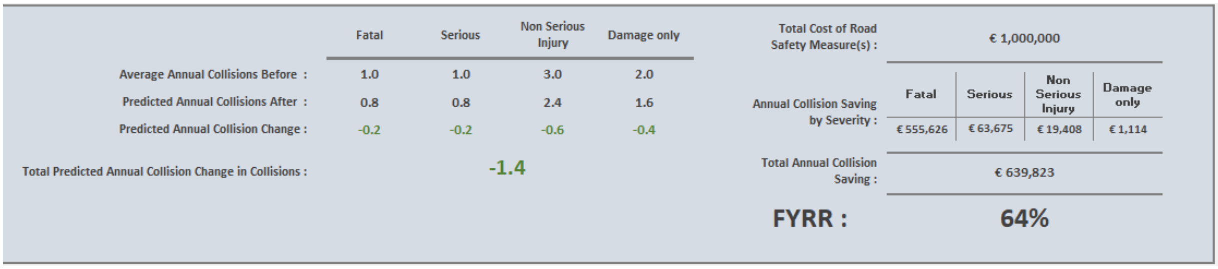


Figure 13: Example screenshot of results section of the tool with sample data

4.2 Irish CMFs included in the tool

Based on the output of the modelling (Section 2.3), Table 22 to Table 25 summarise the Irish CMFs included in the tool. These variables are included as countermeasures for selection in the calculator as described in the tables.

Table 22: Irish CMFs included in the calculator from the Motorway model

Variable	CMF	Interpretation of CMF	Associated countermeasure in the calculator
Gradient	$e^{-0.176} = 0.839$	Decreasing the absolute maximum gradient by 1 degree decreases the number of collisions by 16%.	Decrease in absolute maximum gradient by [1/2/3/4/5] degrees
HGV %	$e^{-0.01804} = 0.982$	Decreasing the proportion of HGVs by 1% decreases the number of collisions by 2%.	Decrease in proportion of HGVs by [1/2/3/4/5]%
Radius	$e^{-0.187} = 0.829$	Increasing the minimum radius by 1000m decreases the number of collisions by 17%.	Increase in minimum radius by [1000/2000/3000]m

[12] The three methods for combining the effect of countermeasures used in the tool are the ‘multiplicative’ (Independent effects) method, the ‘dominant common residuals (DCR)’ method and the ‘minimum CMF’ (dominant effects) method. See ‘An exploratory analysis of models for estimating the combined effects of road safety measures’ (Elvik, 2009) for a clear and detailed description of these methods and when they are most appropriate.

Table 23: Irish CMFs included in the calculator from the Dual Carriageway model (note, major junctions excluded from this table as this coefficient was not in the expected direction – see Section 2.3.2.2 for a full discussion on this)

Variable	CMF	Interpretation of CMF	Associated countermeasure in the calculator
Median barrier	$e^{-0.0102} = 0.990$	Increasing the median barrier proportion by 1% decreases the number of collisions by 1%.	Increase median barrier proportion by [1/2/3/4/5/6/7/8/9/10]%
Radius	$e^{-0.697} = 0.498$	Increasing the minimum radius by 1000m decreases the number of collisions by 50%.	Increase in minimum radius by [1000/2000/3000]m
Commercial access	$e^{-0.019} = 0.981$	Decreasing the number of commercial access points by 1 per km decreases the collision number by 2%.	Decrease number of commercial accesses per km by [1/2/3]

Table 24: Irish CMFs included in the calculator from the Single Carriageway model

Variable	CMF	Interpretation of CMF	Associated countermeasure in the calculator
Gradient	$e^{-0.169} = 0.845$	Decreasing the absolute maximum gradient by 1 degree decreases the number of collisions by 16%.	Decrease in absolute maximum gradient by [1/2/3/4/5] degrees
Minor junctions	$e^{-0.132} = 0.876$	Decreasing the number of minor junctions per km by 1 decreases the collision number by 12%.	Decrease number of minor junctions per km by [1/2/3]
Radius	$e^{-0.073} = 0.930$	Increasing the minimum radius by 1000m decreases the collision number by 7%.	Increase in minimum radius by [1000/2000/3000]m
Commercial access	$e^{-0.015} = 0.985$	Decreasing the number of commercial access points by 1 per km decreases the collision number by 1%.	Decrease number of commercial accesses per km by [1/2/3]
CSC % (skid)	$e^{-0.00186} = 0.998$	Increasing the proportion of road with CSC % above the threshold by 1 decreases the collision risk by 0.2%.	Resurface a road of which [25/50/75/100]% was below the skid resistance threshold

Table 25: Irish CMFs included in the calculator from the Legacy road model

Variable	CMF	Interpretation of CMF	Associated countermeasure in the calculator
Gradient	$e^{-0.054} = 0.947$	Decreasing the absolute maximum gradient by 1 degree decreases the number of collisions by 5%.	Decrease in absolute maximum gradient by [1/2/3/4/5] degrees
Minor junctions	$e^{-0.081} = 0.922$	Decreasing the number of minor junctions per km by 1 decreases the collision number by 8%.	Decrease number of minor junctions per km by [1/2/3]
Radius	$e^{-0.052} = 0.949$	Increasing the minimum radius by 1000m decreases the collision number by 5%.	Increase in minimum radius by [1000/2000/3000]m
Commercial access	$e^{-0.020} = 0.980$	Decreasing the number of commercial access points by 1 per km decreases the collision number by 2%.	Decrease number of commercial accesses per km by [1/2/3]
CSC % (skid)	$e^{-0.00298} = 0.997$	Increasing the proportion of road with CSC % above the threshold by 1 decreases the collision risk by 0.3%.	Resurface a road of which [25/50/75/100]% was below the skid resistance threshold

4.3 Future potential improvements to the tool

The tool could be developed further in the future by:

- **Breaking down countermeasure effectiveness by different collision types**. This relies on more data on collision types being available. In the first instance, this could utilise the collision type information (where it is available) in the Clearinghouse data.
- **Allowing the user to specify the amount of overlap between countermeasures to give more accurate combined reductions.** At present the tool presents an overall predicted collision reduction on the main calculator page, with an optimistic and pessimistic bound on the 'Calculations with CMFs' page. If the user could input knowledge of the overlap between countermeasures, the tool could give a more accurate estimate (for example, by picking the lower bound reduction if overlap was known to be strong).
- **Links to other data sources** (for example, data repositories similar to Clearinghouse) with more countermeasures. If the accident prediction models are improved, further CMFs could be added from these too.

- **Countermeasures could be categorised in the tool by relevance or applicability to Irish roads.** At present, general filters are applied to the CMFs from Clearinghouse to eliminate those that are not applicable (for example removing CMFs that apply to roads with a larger number of lanes than seen on TII roads). An individual assessment of countermeasures or countermeasure groups could highlight those that are particularly useful for Irish roads.

5 Conclusions

Relative to other European countries, Ireland is a small country with low flows and few road collisions. This can present methodological challenges with using Generalised Linear Modelling to develop APMs. The zero-inflated models used in this project, however, have overcome these challenges and demonstrate that the outcomes of these models can inform road safety intervention. In particular, the models show that to reduce collision risk:

- Reducing the number, or improving the safety of, minor junctions and access points onto the network could reduce collision risk.
- On dual carriageways, increase the proportion of median barriers decreases the risk on a segment.
- It is important to ensure the skid resistance (CSC %) meets the defined minimum thresholds on single and legacy roads.
- The geometry of the road influences collision risk: gradient and radius were common significant predictors of collision risk across all models.

APMs have been found to be a more cost-efficient way to carry out evaluation of safety interventions compared to studies of individual features or sites. This research highlighted limitations associated with using existing data sets collected for other purposes. Some improvements have been made to the TII datasets (detailed in Appendix B) but some challenges remain. Suggestions for future improvements to the models have been made in Section 2.4 and for the associated collision reduction calculator in Section 4.3.

The calculator developed in this project can be used by regional and local road safety engineers to support them in identifying effective road safety interventions, understanding the potential collision reductions which could be achieved with these, and feed into the (FYRR) economic appraisal of the measures.

References

Chowdhury, S., Makosa, H., Harpham, N., Collis, C., Wallbank, C., & Fletcher, J. (2022). *TII268 Lot1 Collision Prediction Model for the Irish National Road Network - Interim Report.* Wokingham: TRL Limited.

Google. (2023a). *Google Maps*. Retrieved June 2023, from https://www.google.com/maps/@53.7268466,-7.8006122,15.25z

Google. (2023b). *Google Maps*. Retrieved June 2023, from https://www.google.com/maps/@53.9722464,-8.3032212,15z

Google. (2023c). *Google Maps*. Retrieved June 2023, from https://www.google.com/maps/@52.4484007,-9.4858798,15z

Hensher, D. A., & Stopher, P. R. (1979). *Behavioural Travel Modelling.* Routledge Library Edtions: Global Transport Planning.

Acronyms

AADT	Annual Average Daily Traffic
AIC	Akaike's Information Criterion
APM	Accident Prediction Model
BIC	Bayesian Information Criterion
CMF	Crash Modification Factors
CSC	Characteristic SCRIM (skid resistance) Coefficient
FYRR	First Year Rate of Return
GIS	Geographic Information System
GLM	Generalised Linear Modelling
GPS	Global Positioning System
HGV	Heavy Goods Vehicle
KSI	Killed or Seriously Injured
MAD	Mean Absolute Deviance
MSPE	Mean Squared Prediction Error
NRDO	National Roads Design Offices
NTpM	National transport model
PCA	Principal Component Analysis
PMS	Pavement Management Survey
PRACT	Predicting Road Accidents – a Transferable methodology across Europe
RRSE	Regional Road Safety Engineers
SCRIM	Sideway-force Coefficient Routine Investigation Machine (Skid resistance)
TII	Transport Ireland Infrastructure
TRL	Transport Research Laboratory
VRS	Vehicle Restraint Systems

Appendix A Variables not included in the models

This appendix outlines the variables that were not possible to include in the modelling and the reasons for these.

<table>
<tr><th>Models</th><th>Variable</th><th>Description</th><th>Reason this was not used in the modelling</th></tr>
<tr><td>All</td><td>Percentage rear end</td><td>Percentage of collisions that are rear-end</td><td>The response variable in the models was number of collisions and thus it does not make sense that the percentage of collisions which are rear end collisions would be a predictor variable for this measure.</td></tr>
<tr><td>All</td><td>LV AM peak speed</td><td>Modelled light vehicle (LV) peak speed in the morning</td><td rowspan="4">Initial modelling carried out with the speed variables indicated that the coefficient for these data was negative. This was unexpected since a negative coefficient indicates that if speed increases, collision risk decreases; which is contrary to well-researched evidence from literature showing that increasing speeds results in an increased collision risk.

The speed data used for the modelling was provided from the National Transport Model (NTpM). This is used for understanding delay and due to the way in which it is calculated (including being capped based on a maximum allowable speed function), this does not accurately represent the actual traffic speeds, and therefore collision risk, on a given link.

Alternative sources of speed data were (re)considered but none of these were identified as a practical route to obtain the speed data needed for the modelling. As a result, a decision was made to develop the models without speed.</td></tr>
<tr><td>All</td><td>HV AM peak speed</td><td>Modelled heavy vehicle (HV) peak speed in the morning</td></tr>
<tr><td>All</td><td>LV inter-peak speed</td><td>Modelled light vehicle (LV) inter-peak speed</td></tr>
<tr><td>All</td><td>HV inter-peak speed</td><td>Modelled heavy vehicle (HV) inter-peak speed</td></tr>
<tr><td>All</td><td>Risk rating</td><td>Highest risk rating for the segment</td><td>This variable could not be included in the modelling due to the large volume of missing data. The measurement points for this data are sparse, and with a 50m buffer applied to the network, many of the segments had no risk rating applied to them.</td></tr>
</table>

Models	Variable	Description	Reason this was not used in the modelling
Dual, single, legacy	Median width and hard shoulder width	Median width and hard shoulder width categorisations.	These variables were collected manually using a visual inspection of Google Earth for motorways. However, due to the scale of the data collection requirements, it was not deemed possible to carry out the same inspection for the dual, single and legacy networks; as a result, these variables are not included in the models.
Single, legacy	Verge barrier %	Percentage of the segment with a verge barrier on the nearside.	This variable was not included in the single and legacy models as the data provided by TII were unreliable, and the task to collect robust data by a visual inspection was deemed too great. TII are currently collecting more reliable data but it was not available within the timeframes of this project.
Motorway	Access Business/ Commercial/ Residential per kilometre	Access density to business/ commercial/ residential premises	This variable was calculated for all road types but was not used for the motorway modelling as it was assumed that no accesses of these types would be directly onto the motorway network.

Appendix B Data improvements

During the initial segmentation and modelling activities, several issues with the data were identified which influenced the results of the preliminary models developed. As a result, a series of data improvements were carried out to resolve these before the network segmentation and modelling were updated (the results of which are presented in Section 2.3). This appendix summarises the data improvements made in the course of this work.

As outlined in the interim report (Chowdhury, et al., 2022), the base map layer selected for the analysis was the TII GIS base layer. During the course of the modelling, some inaccuracies in the road type classification were identified. To rectify these, the PRIME 2 data were used to identify any discrepancies between the two datasets, and then used to update the segments which were classified incorrectly.

Using the TII GIS layer to calculate the junction counts on each segment led to substantial inaccuracies in this count: there were large numbers of minor junctions missing from the shapefile. The PRIME data was again used to improve the accuracy of these counts. Some manipulation was needed for this:

- Roundabouts in the PRIME2 data often had multiple points recorded for each junction. To ensure these weren't double counted, a method for combining these points was used.
- Slip road counts were inconsistent: some slip roads recorded a central point to cover both the on and off junctions on each carriageway, others had multiple points. To ensure reliable junction counts for each segment, the line data from PRIME2 was used, converted into a centralised point and then matched the closest "main road" line.
- A suitable buffer was defined to capture junctions located at the end of each segment. This ensured that collisions occurred on these junctions were explained by the junction density variables associated with that segment.

The datasets provided by TII did not provide accurate information on median or hard shoulder width, or the presence of barriers. Some data is collected, and activities are underway to increase the collection of this information; however, this was not available during the timeframes of this project. As a result, some limited manual data collection was carried out using Google Earth imagery to supplement the motorway and dual carriageway datasets. Specifically, information on the following road characteristics were collected at 1km intervals, combined with existing data where this was available, and used to estimate the median and hard shoulder characteristics:

- Median width (motorways only)
- Hard shoulder width (motorways only)
- Median barrier presence (motorways and dual carriageways)
- Verge barrier presence (motorways and dual carriageways)

Appendix C KSI models

The killed and seriously injured (KSI) models presented in this appendix were created using the same methodology as outlined in Section 2.3, but the response variable used was KSI collisions only, rather than all reported collisions. Since KSI collisions are relatively uncommon compared to slight or damage only collisions, there are substantially more segments[13] with zero collisions recorded. As a result, the models created are more uncertain than those for all collisions. This is demonstrated by the typically lower McFadden's R-squared values (a measure of model fit) for each road type (the exception of this is dual carriageways for which the KSI model appears to better fit these data).

As with the 'all collisions' models, some caution should be taken when interpreting the results of the KSI models since it was not possible to include all the variables which are known to influence collision risk (e.g. speed). In addition, due to the smaller number of segments with non-zero collision numbers, the models are less robust: some coefficients (and standard errors) cannot be estimated or are very large, even using the zero-inflated models.

Despite these limitations, it is interesting to compare significant variables for these models with those for all collisions; in many cases the variables which are significant predictors of KSI models are a smaller subset of those which were found as predictors for the all collisions models, suggesting that use of the latter models (which fit the data much better) in the subsequent tool development is still valid if the focus is on reducing the more severe collisions.

C.1 Motorway

The base model included AADT and segment length variables, both in power form. The likelihood ratio test suggests a zero-inflated Poisson model is appropriate.

The stepwise procedure was followed for this model but the base model (which includes only segment length and AADT) was found to be the best fitting model – see Table 26.

Table 26: Variables included in motorway model for KSI collisions (variables of interest in black, others also included in the model in grey)

	Variable	Coefficient	p-value
Count model	*Intercept*	*7.247*	*p<0.001*
	Log(segment length)	0.809	p<0.001
	Log(AADT)	0.601	p<0.001
Zero inflated model	*Intercept*	*20.967*	*0.30 (ns)*
	Log(AADT)	-2.594	0.27 (ns)

[13] The segmentation used for these models was the same as that reported in the earlier sections of this report.

For the significant variables:

- The segment length coefficient is very similar to the 'all collisions' model (0.809 compared with 0.765).
- The AADT variable is quite different however: 0.601 for KSI collisions (suggesting KSI collisions increase as AADT increases but at a decreasing rate) and 1.157 for all collisions (suggesting all collisions increase as AADT increases but at an increasing rate).

The McFadden's R-squared value was 0.10 suggesting this KSI model is not particularly strong.

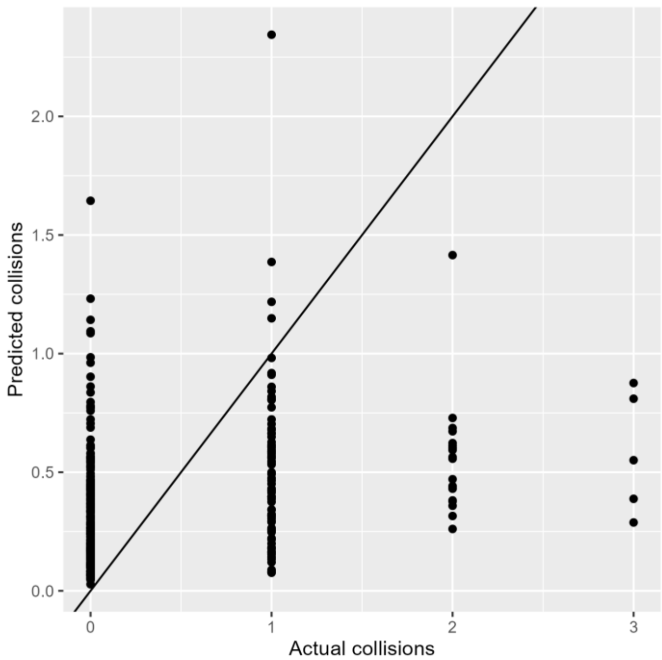


Figure 14: Assessment of model predictions against actual collision numbers for the motorway KSI collisions model

Table 27: Prediction accuracy for the motorway model for KSI collisions

	Final model
Mean (MAD)	0.45
Sqrt(mean (MSPE))	0.62

C.2 Dual carriageway

The base model included AADT and segment length variables, both in power form.

The likelihood ratio test result indicated that the zero-inflated negative binomial model had significantly better goodness-of-fit than the equivalent Poisson model with $p<0.01$.

Table 28 shows the model coefficients and Table 29 assesses the model fit.

Table 28: Variables included in dual carriageway model for KSI collisions (variables of interest in black, others also included in the model in grey)

	Variable	Coefficient	p-value
Count model	*Intercept*	*-11.608*	*p<0.001*
	Log(segment length)	1.105	p<0.001
	Log(AADT)	1.187	p<0.001
	Minor junctions	0.075	0.03
	Median barrier %	-1.392	p<0.001
	Log(theta)	*6.311*	*0.76 (ns)*
Zero inflated model	*Intercept*	*-11.194*	*0.28 (ns)*
	Log(AADT)	1.007	0.31 (ns)

For the significant variables:

- The segment length coefficient is quite different to the coefficient for all collisions: 1.105 for KSI collisions (suggesting KSI collisions increase as length increases but at an increasing rate) and 0.597 for all collisions (suggesting all collisions increase as length increases but at a decreasing rate).
- The AADT variable is very similar to the 'all collisions' model (1.187 compared with 1.144).
- The median barrier % variable is also included in the 'all collisions' model (Section 2.3.2.2) with a similar magnitude for the coefficient.
- Minor junction density was not included in the 'all collisions' model, however (major junction density did feature), suggesting that the minor junctions might be more influential for KSI collision risk.

McFadden's R-squared for this model (0.21) is good suggesting this model performs quite well, and better than the 'all collisions' model which only scored 0.13 on this measure.

Table 29: Goodness of fit measures for the dual carriageway model for KSI collisions

	Base model	Final model
AIC	307	266
BIC	326	292
McFadden R-Squared	0.07	0.21

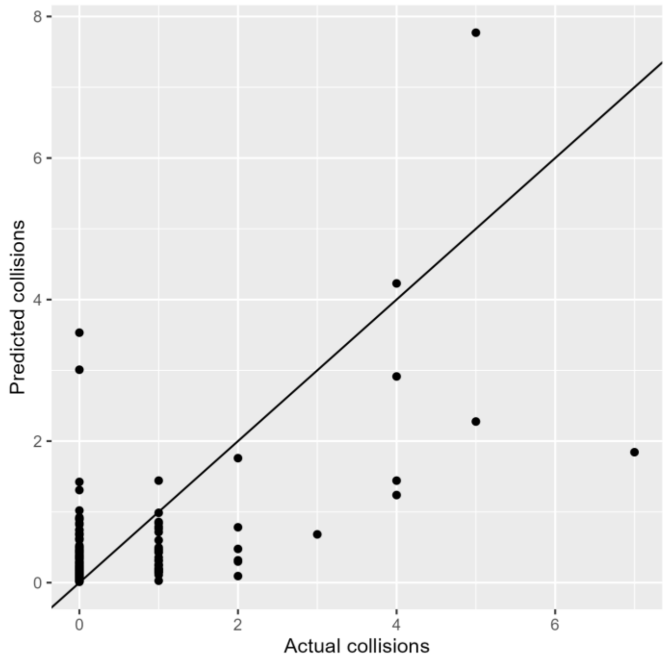


Figure 15: Assessment of model predictions against actual collision numbers for the dual carriageway KSI collisions model

Table 30: Prediction accuracy for the dual carriageway model for KSI collisions

	Final model
Mean (MAD)	0.48
Sqrt(mean (MSPE))	0.83

C.3 Single carriageway

The base model included AADT and segment length variables, both in power form.

The likelihood ratio test result indicated that the zero-inflated negative binomial model had significantly better goodness-of-fit than the equivalent Poisson model with $p<0.01$.

Table 31 shows the model coefficients and Table 32 assesses the model fit.

Table 31: Variables included in single carriageway model for KSI collisions (variables of interest in black, others also included in the model in grey)

	Variable	Coefficient	p-value
Count model	*Intercept*	*-7.758*	*p<0.001*
	Log(segment length)	0.898	p<0.001
	Log(AADT)	0.720	p<0.001
	Minor junctions	0.080	p<0.001
	Crossfall	0.373	p<0.001
	Log(theta)	*1.046*	*p<0.001*
Zero inflated model	*Intercept*	*5.012*	*0.98 (ns)*
	Log(AADT)	-1.682	0.95 (ns)

For the significant variables:

- The segment length variable is very similar to the 'all collisions' model (0.898 compared with 0.841).
- The AADT variable is very similar to the 'all collisions' model (0.720 compared with 0.877).
- The minor junction density variable is also included in the 'all collisions' model (Section 2.3.3.2) with a similar magnitude for the coefficient.
- Crossfall (i.e. transverse fall) was not, however, included in the 'all collision' model (other physical segment characteristics including gradient and radius did feature), suggesting that the crossfall might be more influential for KSI collision risk.

The McFadden's R-squared value was 0.11 suggesting this model isn't particularly strong.

Table 32: Goodness of fit measures for the single carriageway model for KSI collisions

	Base model	Final model
AIC	3230	3191
BIC	3265	3237
McFadden R-Squared	0.10	0.11

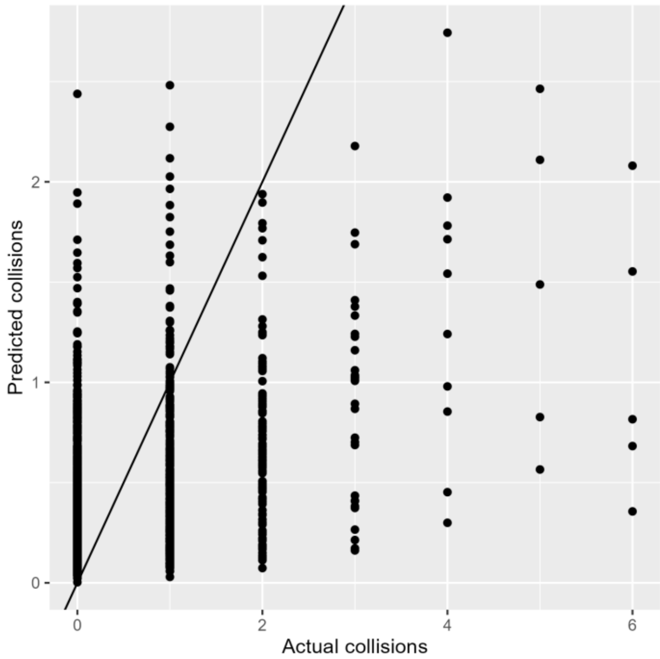


Figure 16: Assessment of model predictions against actual collision numbers for the single carriageway KSI collisions model

Table 33: Prediction accuracy for the single carriageway model for KSI collisions

	Final model
Mean (MAD)	0.45
Sqrt(mean (MSPE))	0.68

C.4 Legacy roads

The base model included AADT and segment length variables, both in power form. The likelihood ratio test suggests a zero-inflated Poisson model is appropriate.

The stepwise procedure was followed for this model but the base model (which includes only segment length and AADT) was found to be the best fitting model – see Table 34.

Table 34: Variables included in legacy road model for KSI collisions (variables of interest in black, others also included in the model in grey)

	Variable	Coefficient	p-value
Count model	*Intercept*	*-5.631*	*p<0.001*
	Log(segment length)	0.798	p<0.001
	Log(AADT)	0.541	p<0.001
Zero inflated model	*Intercept*	*-2.701*	*0.86 (ns)*
	Log(AADT)	-0.185	0.92 (ns)

For the significant variables:

- The segment length variable is very similar to the 'all collisions' model (0.798 compared with 0.970).
- The AADT variable is very similar to the 'all collisions' model (0.541 compared with 0.680).

The McFadden's R-squared value was 0.09 suggesting this model is not particularly strong.

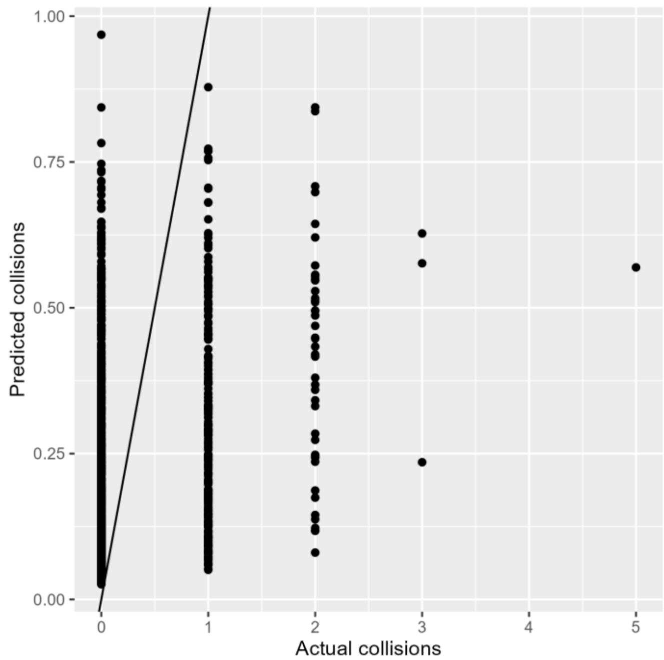


Figure 17: Assessment of model predictions against actual collision numbers for the legacy road KSI collisions model

Table 35: Prediction accuracy for the legacy road model for KSI collisions

	Final model
Mean (MAD)	0.31
Sqrt(mean (MSPE))	0.46

Appendix D Online survey responses

TII 268 Research Programme - APM

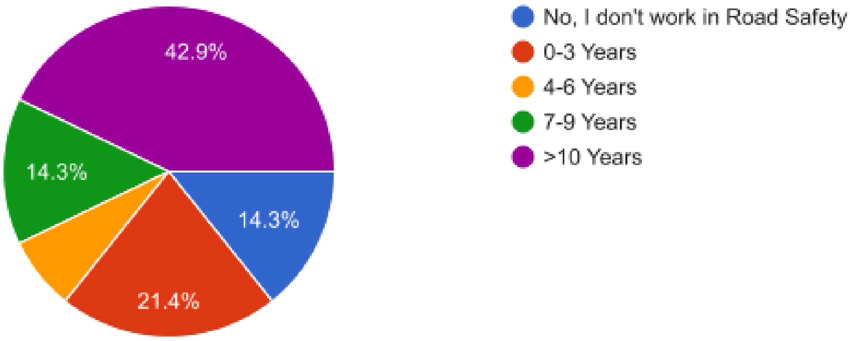
14 responses

Publish analytics

Do you currently work in Road Safety? If Yes, for how many years?

Copy

14 responses



What data is collected, and how is it collected, for Road Safety Improvement Schemes, for example, collision records, local knowledge, site visits, collision investigations, etc.?

14 responses

Collision Records

Collision history
No RSA data anymore
Locals

All of the above
Desk top study
Google street view

Collision records where available
Local knowledge inl. Area Engineer
site visits

Detailed Collision data for each collision on network. All incidents on the network are also included which gives a good network knowledge

Collision records from RSA. Discuss with Area Engineer. Elected members can provide / acquire valuable information. Site visit essential

Site visits/ photos/ videos
Trying to obtain Historic RS data from RSA at moment and deal with GDPR issue

Collision Records - RSA Database, Gardai requests, Local authority Feedback - Area Engineer, Local Knowledge, Site Visits

Road Safety Authority accident stats. TII stats from Lucy Curtis. RSI items from Liz Kennedy. Local Knowledge from residents, area staff and LA employees who know the area. LA temp traffic count including speed. Relevant TII Counters . Site visits and surveying (Topo etc,)

Collision history - map road/ RSA
Local Gardai
Loical residents (informal)

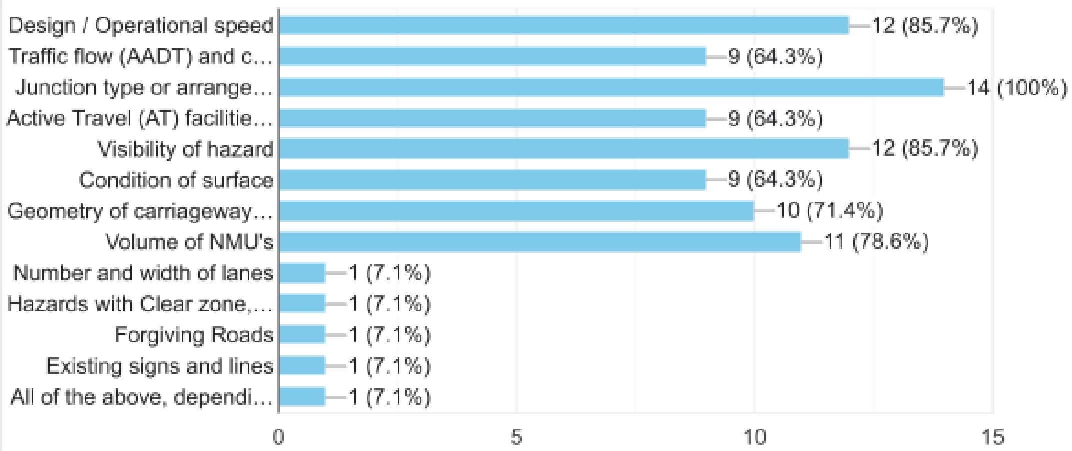
All those mentioned. We look at the collision narratives from the pulse system. We gather existing site information, using one or more of topo surveys, cameras, drones potentially.

As a Road Safety Inspection Engineer (RSIE) and new to the role this year, my focus is Road Safety Inspections (RSIs) which are a pro-active approach to identifying safety issues before a problem manifests itself. A team of at least 2 people drive and video each national route plus 200m up each side road in each direction during the hours of daylight and darkness to identify any potential safety issues which are compliled in a report with spreadsheets of tagged items. These tagged items are added to ArcGIS for each county. For a proposed road safety improvement scheme, a feasability and options report is initiated by the local authority (LA) based on RSIs, collision records, site visit and local knowledge. RRSE/RSIE liaise with LA and provide input.

Collision records ,Local knowledge & site visits

What are the typical key inputs of the existing carriageway for calculating the expected road safety benefits?

14 responses



How are countermeasure packages (schemes) developed?

14 responses

Depending on collision

LA Liason with TII

Desktop review, site visit, F x O report, concept design which may include options, cost estimate, recommendation, funding approval, detailed design

Assessment of identifible problems
Assess collision modification factors
Use engineers experience
Produce Design Options

Local needs, future design and existing issues

N/a

Generally primary schemes that are on cycle connects/ pathfinder and also if there is a high volume of peds/ cyclists that warrant better active travel infrastructure

Based on collisions,at any available records. Type of collisions and relation to road parameters/ properties

HD15, HD17, HD28 schemes or minor or major schemes developed following discussions with Regional Road Safety Inspector and TII Inspector. Developed as per appropriate TII standard (egg. PMA for minor and major schemes)

1) Determine main aspects of the road that contributed to the collision
2) look to improve them by utilising the effective solutions that will reduce future collisions and be within budget / FYRR use CMF

A lot comes from experience. The main thing is to try to establish what is happening and what factors are contributing. We would then assist in coming up with countermeasures that might address what's going wrong. A number of options will be compared in terms of collision reduction and cost (a First Year Rate of Return (FYRR) will generally be calculated).

Once a preferred option has been decided/approved (F&O report), the local authority or a consultant develop a solution with input from the RRSE/RSIE. Solutions will depend on issue and location.

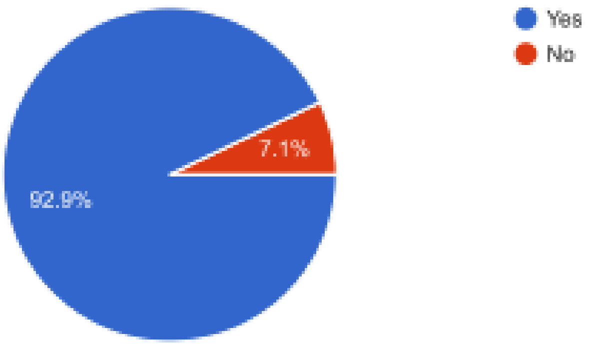
Through design

Looking at collsions and reduction- care not to add to problem- clear zone

Are safety benefits calculated, for example expected reduction in collisions, injuries, or severity etc.? Yes / No / Other (please specify).

Copy

14 responses



How do you assess the effectiveness of the safety benefits for the proposed countermeasures, for example, potential reduction in collisions, injuries, or severity etc.?

14 responses

Before and after analysis

Reduction in collisions

Potential design

Reduction in collisions - only

Reduction in collisions
Severity of injuries
Monitor over periods of time to react

Collision Modification Factors
FYRR

Past Monitors is usually assess quality of build

Recording at speed data, collision records

Stage 3 or Stage 4 RSA. Observations on site. Lack of evidence of collisions such as damaged signage and verges. No reports of collisions. Local perceptions and asking has there been an improvement,any near misses etc. (ask residents and area staff)

Collisions 3 years before V 3 years after

We have a FYRR spreadsheet that calculates the return, based on collision costs, scheme costs and estimated collision savings. Where we don't have collision information, crash modification factors (CMFs) can be used.

Collision Modificatoin Factor (CMF)

CMF

Reduction

What are the practical challenges to calculating expected reduction in collisions, injuries or severities etc.?

14 responses

Collision data delay
Appropriate collision details

Information

Suitable collision modification factors

Ability to find suitable CMF's

Sharing of information across parties
Different outcomes/ Solutions for a scheme
Information (Real Time) of driver behavior especially during changes in weather i.e. hail/ heavy rain driver changing/ not changing their driving style

Finding CMF for similar scheme as design proposed. Most CMFs are based on stats in different jurisdictions with different driver behavior. Accessibility of previous collision data, particularly details of how collisions occurred

Don't do this usually

Access to data, due to GDPR collecting the data

Have not done this recently. This process needs to be simple and acknowledge there are unknowns and result is in no way definitive (its an estimate)

Look at past monitoring
Delay in getting collision data
Lack of details on collision esp. minor injuries and major damage

At the end of the day, you are guessing - but it is based on experience.

Many variables which can be difficult to predict or measure.
Any change to location can effect driver/VRU behaviour.

AVAILABILITY OF DATA

Available data Experience

What are the scheme requirements to justify commencement of road safety improvements?

14 responses

N/a

Political buy in

Known safety issues

N/A Road Authority Fuction

Reduce collisions
Public view/ perceptions
Reduction of maintenance / safer working for maintenance teams

Identified road safety issues and collision history

VRS are helping more vulnerable in urban areas hence a scheme is needed

HCL's identification of collision clusters

Cost Benefit Analysis - reductions in collisions from an economic stand point

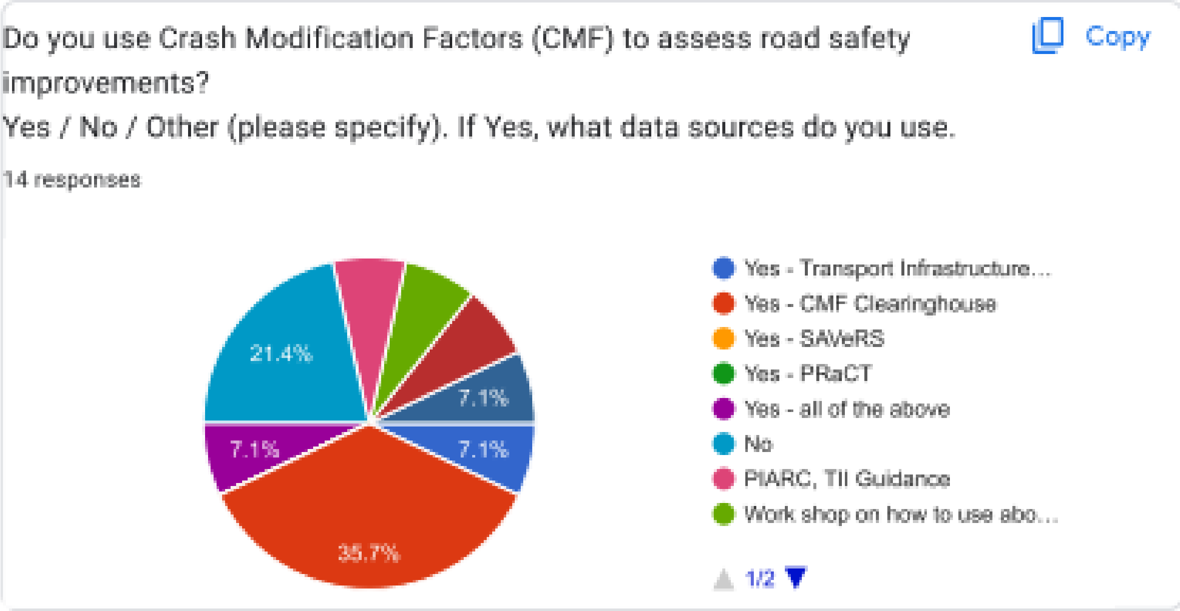
F x O report submitted and get approved

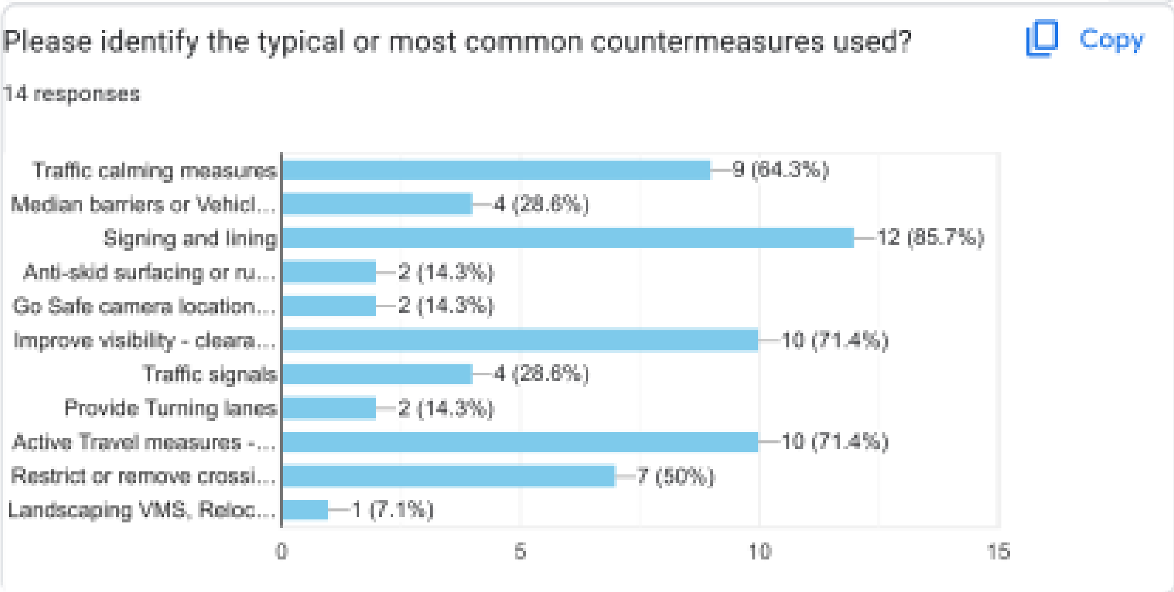
This is set out in the RSIS approval procedure. Generally, sites are identified as a HCL or RSI under their respective TII programmes. LAs can also nominate sites with our agreement.

Can include: Identified RSIs (Road Safety Inspection tags). High risk of Collision. Poor VRU provisions especially where high volume of VRUs. Speeding.

Budget, FYRR,

Aim is always to get improvement- focus on legacy road is consistent works


Do you use Crash Modification Factors (CMF) to assess road safety improvements?
Yes / No / Other (please specify). If Yes, what data sources do you use.
14 responses
21.4%
7.1%
7.1%
7.1%
35.7%
Yes - Transport Infrastructure...
Yes - CMF Clearinghouse
Yes - SAVeRS
Yes - PRaCT
Yes - all of the above
No
PIARC, TII Guidance
Work shop on how to use abo...
1/2


Please identify the typical or most common countermeasures used?
14 responses
Traffic calming measures
9 (64.3%)
Median barriers or Vehicl...
4 (28.6%)
Signing and lining
12 (85.7%)
Anti-skid surfacing or ru...
2 (14.3%)
Go Safe camera location...
2 (14.3%)
Improve visibility - cleara...
10 (71.4%)
Traffic signals
4 (28.6%)
Provide Turning lanes
2 (14.3%)
Active Travel measures -...
10 (71.4%)
Restrict or remove crossi...
7 (50%)
Landscaping VMS, Reloc...
1 (7.1%)
0
5
10
15

Do you have anything further to add?

7 responses

N/a

GDPR is hindering access to collision data and engineering solution development

safer working area/ reduction of need to be "on road" for maintenance teams should be included more
Increased driver/ user education on reason for scheme/ awareness of all road users/ others
Increased driver education on changing behaviors of driving for weather/ works etc.

Work in Active Travel (Urban Areas) so not directly involved but I do consider road safety aspects on a daily basis

Full time Road Safety Audit Team leader approx 1/5 of time Auditing 6 or 7 yrs.

Telemetrics could potentially provide more data, especially on driver behaviour.

It is important to widen the data base. Added value would be to use data from Transport for Scotland and Road Service NI

Appendix E Technical Process Flow

E.1 Filters applied before the user interacts with the tool

Only the relevant CMFs for Irish roads are incorporated into the tool from Clearinghouse. The filters that determine these CMFs are as follows (with field names from the Clearinghouse data given in italics):

- Star rating at least 3 (*qualRating* = 3,4,5)
- *CrashType* = 'All', 'Not specified' or 'Day time, Nighttime'. (this removes crash types that are definitely not all crashes).
- Time of day not specific to day or night (*crashTOD*= 'All', '(Blank)', 'Not specified').
- Minimum number of lanes (*minNumLanes*) no more than 4 if one direction, and no more than 8 if both directions and no more than 8 if number of directions is not specified. The field *numLanesDirection* specifies the number of directions.

E.2 Steps for the user interacting with the tool

(Note that the tool predominantly works with 'collision change percentages' rather than 'CMFs' when displaying collision reductions. It is more convenient to work with 'CMFs' when describing the process flow in technical detail here.)

1. The user selects the road type that they are interested in from four options:
    a. Motorway
    b. Dual carriageway]
    c. Single Carriageway
    d. Legacy

Note that the tool only applies to TII managed roads.

2. The user inputs the number of collisions, by severity, at that site, and the number of years the data is from. The tool then calculates the yearly number of collisions of each severity at the site.

3. Once the user has selected the road type this filters down the applicable CMFs (these are presented as 'collision reduction %s' not 'CMFs'):
    a. Only the CMFs from our models that relate to that road type are shown
    b. Only the CMFs from Clearinghouse that relate to this road type are shown

The filters applied to Clearinghouse for each road type are:

- Motorway:
    - Minimum number of lanes not more than 4 if one direction and not more than 8 if both directions or if number of directions not specified

   - Remove CMFs where the maximum speed limit (*maxSpeedLimit*) is less than 100kph or where the minimum speed limit (*minSpeedLimit*) is greater than 120kph
   - Road division type (*roadDivType*) not 'undivided'
   - Road type not 'Local'
- Dual:
   - Minimum number of lanes not more than 3 if one direction and not more than 6 if both directions or if number of directions not specified
   - Remove CMFs where the maximum speed limit (*maxSpeedLimit*) is less than 80kph or where the minimum speed limit (*minSpeedLimit*) is greater than 120kph
   - Road division type (*roadDivType*) not 'undivided'
   - Road type not 'Local'
- Single:
   - Minimum number of lanes not more than 2 if one direction and not more than 4 if both directions or if number of directions not specified
   - Remove CMFs where the maximum speed limit (*maxSpeedLimit*) is less than 50kph or where the minimum speed limit (*minSpeedLimit*) is greater than 100kph
   - Road division type (*roadDivType*) = 'All' or 'Undivided' or not specified
- Legacy:
   - Minimum number of lanes not more than 2 if one direction and not more than 4 if both directions or if number of directions not specified
   - Remove CMFs where the maximum speed limit (*maxSpeedLimit*) is less than 50kph or where the minimum speed limit (*minSpeedLimit*) is greater than 100kph
   - Road division type (*roadDivType*) = 'All' or 'Undivided' or not specified

4. The user then selects the countermeasures of interest and the tool gives a CMF for each countermeasure, by severity. When the countermeasures are from the Clearinghouse:
   a. If there are multiple CMFs for a chosen countermeasure, use the CMF that applies to 'all' crash severities (defined below*) and allocate this to each of the four severity types in the tool. If there are multiple countermeasures applying to 'all' crash severities (e.g. from different studies), average the CMFs to give an overall CMF for each of the severity types.
   b. If there are no CMFs for that countermeasure that apply to 'all' crash severities, for each severity take all the CMFs in the Clearinghouse database

that apply to that severity and average these. In this case the CMF presented in the tool for different severity types may not be the same.

*The field in Clearinghouse specifying crash severity is *crashSeverityKABCO*. Where the value of this field is 'All', 'Blank', 'Not specified' or 'K,A,B,C,O' we treat the CMF as applying to 'all' crash severities.

When the user has selected their countermeasures and the associated CMFs have been determined using the method in 4a or 4b above, a table will be presented as in Table 36.

Table 36: Example of CMF presentation by countermeasure (A, B and C) and injury outcome

Countermeasure	Fatal CMF	Serious CMF	Non-serious inj. CMF	Damage only CMF
A	0.9	0.9	0.9	0.9
B	0.8	0.8	0.8	0.75
C	0.7			

In the example here, countermeasure A has a CMF of 0.9 for each of the severity types, countermeasure B has a CMF of 0.8 applying to all injury severities ('KABC'), and a CMF of 0.75 for damage only collisions, and countermeasure C has a CMF of 0.7 only applying to fatal collisions.

5. For each severity (each column above) the CMFs need to be combined. To do this, for each column above, an upper and lower bound is calculated. Assuming all the CMFs are greater than zero and not more than 1:
    a. Lower bound (lower % reduction, the 'pessimistic' option) = dominant common residual (DCR) value. The DCR is calculated by multiplying the CMFs together and raising to the power of the numerically smallest one. Example: for the fatal column above the DCR is (0.7x0.8x0.9)^0.7 = **0.62**
    b. Upper bound (greater % reduction, the 'optimistic' option) = product of the CMFs.

The tool checks that product < DCR < minimum CMF in order for the DCR to be the lower bound. If this is not true, the minimum CMF is the lower bound. Example: For CMFs 0.5 and 0.9, the product is 0.45, the DCR is 0.67 and the minimum is 0.5. So, we use the minimum CMF (0.5) here instead of the DCR.

If any of the CMFs are bigger than 1:

- For the pessimistic bound, take all the CMFs that are less than 1 and combine these as in the lower bound calculation above (which results in DCR if product < DCR < min CMF; min CMF otherwise). Then, multiply this result by all the CMFs that are greater than 1. If there are no CMFs less than 1, the pessimistic result will just be the product of the CMFs bigger than 1. Example for CMFs 0.7, 0.8, 0.9, 1.1 and 1.3: Start by calculating the DCR as (0.7x0.8x0.9)^0.7 = 0.62. This is between the product (0.504) and the minimum (0.7) so we use the DCR. Then multiply 0.62 x 1.1 x 1.3 = **0.89**.

- For the optimistic bound, multiply the CMFs together to give an overall estimated change in collisions. Same example as above bullet: 0.7 x 0.8 x 0.9 x 1.1 x 1.3 = **0.72**.

The tool includes a warning that tells the user they are using a countermeasure that increases collisions if they pick one of these.

6. The upper and lower bounds are averaged to produce an overall estimate to carry through the calculations (see Table 37).

Table 37: Example calculation of average (overall) estimate of the CMFs

Fatal	Serious	Non-serious inj.	Damage only
0.9	0.9	0.9	0.9
0.8	0.8	0.8	0.75
0.7			
Upper = 0.50	Upper = 0.72	Lower = 0.72	Lower = 0.68
Upper = 0.62	Lower = 0.77	Lower = 0.77	Upper = 0.74
Overall = 0.56	**Overall = 0.745**	**Overall = 0.745**	**Overall = 0.71**

7. Using the yearly number of collisions by severity, the tool calculates the estimated number of collisions left after applying the countermeasures, and also the estimated reduction in collisions.
8. To calculate the **FYRR,** the user inputs the value of a collision of each severity type, and the cost of the countermeasures. The tool then computes the benefit (by multiplying the collision values by the yearly reduction in collisions) and the cost (by adding the costs of the individual countermeasures). The FYRR is the benefit divided by the cost, x 100. A FYRR > 100 indicates all the money is made back in the first year.

The main outputs from the tool are:

- % collision reduction, by severity
- Absolute collision reductions yearly, by severity and overall
- First year cost (of countermeasures) and benefit (of saving collisions)
- FYRR (as a %)

Collision Prediction Model for the Irish National Road Network

This report presents the results of Phase 2 of a two-phase project to develop Accident Prediction Models for Transport Ireland Infrastructure. The aim of these models is to assist engineers to better manage the safety of physical road features across its trunk network. Phase 1 reviewed the statistical approaches used by others to develop APMs, reviewed the data available in Ireland to develop these models and made recommendations on how these models could be developed and applied. Phase 2 developed the models and associated practitioners' tools for Ireland.

Other titles from this subject area

PPR2030	TII268 Lot1 Collision Prediction Model for the Irish National Road Network – Phase 1 Report. S Chowdhury, H Makosa, N Harpham, C Collis, C Wallbank & J Fletcher. 2023

TRL
Crowthorne House, Nine Mile Ride,
Wokingham, Berkshire, RG40 3GA,
United Kingdom
T: +44 (0) 1344 773131
F: +44 (0) 1344 770356
E: enquiries@trl.co.uk
W: www.trl.co.uk

DOI: 10.58446/hoqk4513

PPR2031