RSS 2022
8th Road Safety & Simulation International Conference
Road Safety and Digitalization
08-10 June, 2022 • Athens, Greece

A state-of-practice review on Crash Prediction Modelling

Katerina Deliali, PhD
Research Associate

Together with:
Anastasios Dragomanovits, Ourania Basta & George Yannis

National Technical University of Athens
Department of Transportation Planning and Engineering

i-safemodels

International Comparative Analyses of Road Traffic Safety Statistics and Safety Modeling

- **Project Team**
  NTUA Department of Transportation Planning & Engineering (www.nrso.ntua.gr)
  - OSeven Telematics (www.oseven.io)
  - Tongji University (https://en.tongji.edu.cn)
  - Third country partners: University of Central Florida (US), Purdue University (US), Loughborough University (UK), German Aerospace Center (DE)

- **Project Duration**
  36 months (Oct. 2019-Oct. 2022)

European Union
European Regional Development Fund

ΕΣΠΑ 2014-2020
ανάπτυξη - εργασία - αλληλεγγύη

ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ

ΓΕΝΙΚΗ ΓΡΑΜΜΑΤΕΙΑ ΕΡΕΥΝΑΣ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ

EPAnEK 2014–2020
OPERATIONAL PROGRAMME
COMPETITIVENESS • ENTREPRENEURSHIP • INNOVATION

Introduction

- Road crashes:
  - are a public health issue and
  - a leading cause of death in certain age groups and demographics, while
  - fatalities cost most countries 3% of their GDP *(WHO)*.

- Crash occurrence analysis stands as the main approach for the assessment of road infrastructure.

- Crash Prediction Models (CPMs) are the state-of-art in a crash occurrence analysis *(Elvik et al., 2003)*, while they also allow for a proactive (i.e., without relying on crashes) safety assessment of the road.

Methodology (1/2)

- The objective of this study is the review of the state-of-practice in CPM with a focus on practical implications and current trends.

- Database search - Publications:
  - in the English language
  - in scientific journals
  - by associations, organizations, or governments (e.g., Federal Highway Administration)
  - focusing on microscopic level and particularly on road segments
  - focusing on motorized vehicle crashes

Methodology (2/2)

Publications were grouped in the following categories:

- Publications related to AASHTO's Highway Safety Manual Predictive Method *(AASHTO, 2010; 2014)* in addition to the manual itself.
- Publications on the use and development of Crash Modification Factors (CMFs).
- Stand-alone multivariate CPMs are models that include a large number of explanatory variables compared to SPFs, in order to consider site characteristics on their own, without the use of CMFs.
- Machine learning (ML) approaches for CPMs.
- Real-time CPMs.

AASHTO Highway Safety Manual Predictive Method

<u>Base</u> Safety Performance Function (SPF) – *developed for specific road types (e.g., two-lane rural roads)*

$$N_{SPF} = f(AADT, Length)$$

Use of count data models and usually the Negative Binomial Distribution

$$N_{predicted} = N_{SPF} \times (CMF_1 \times CMF_2 \times \ldots \times CMF_y) \times C$$

Where N_{SPF} is the predicted average crash frequency determined for the base conditions according to the SPF_{base};
CMF_1 …. CMF_y are the crash modification factors that account for specific road design and operational characteristics;
C is the calibration factor to adjust the SPF for local conditions related to the network where the model is to be applied;
$N_{predicted}$ is the predicted average crash frequency.

AASHTO Highway Safety Manual Predictive Method

Strengths

- The HSM Predictive Method stands as the most complete approach in crash prediction modeling (state-of-art).
- Covers a wide range of facilities as well as geometric and operational conditions.

Limitations

- Limited transferability of the results; the majority of the States in the US have ended up developing their own crash prediction models instead of calibrating the ones in HSM.
- The process of calibration, i.e., with the use of a single calibration factor is not considered adequate.
- The incorporation of different design and operational elements has received criticism as being simplistic.
- From a practical viewpoint, the method needs advanced expertise in statistical modeling & data

Crash Modification Factors

Method	Strengths	Weaknesses
Before–after with comparison group	Simple and accounts for non–treatment-related time trends.	Difficult to account for regression-to-the-mean.
Before–after with empirical Bayes	Accounts for regression-to-the-mean, traffic volume changes & non–treatment-related time trends.	-Relatively complex. -Cannot include prior knowledge of treatment. -Cannot consider spatial correlation.
Full Bayes	-Reliable results with small sample sizes. -Can include prior knowledge, spatial correlation, and complex model forms in the evaluation process.	Implementation requires a high degree of training.
Cross-sectional	-Possible to develop CMFunctions. -Allows estimation of CMFs when conversions are rare. -Useful for crash prediction.	-CMFs may be inaccurate for a number of reasons: inappropriate functional form, omitted variable bias, correlation among variables.
Case-control	-Useful for studying rare events because the number of cases and controls is predetermined. -Can investigate multiple treatments per sample.	-Can only investigate one outcome per sample. -Does not differentiate between locations with one crash or multiple crashes. -Cannot demonstrate causality.
Cohort	-Useful for studying rare treatments because the sample is selected based on treatment status. -Can demonstrate causality.	-Only analyzes the time to the first crash. -Large samples are often required.
Meta-analysis	-Can be used to develop CMFs when data are not available for recent installations, and it is not feasible to install the strategy and collect data. -Can combine knowledge from several jurisdictions and studies.	-Requires the identification of previous studies -Requires a formal statistical process. -All studies included should be similar in terms of data used, outcome measure, and study methodology.
Expert panel	-Can be used to develop CMFs when data are not available for recent installations, and it is not feasible to install the strategy and collect data. -Can combine knowledge from several jurisdictions and studies. -Does not require a formal statistical process.	-Traditional expert panels do not systematically derive precision estimates of a CMF. -Possible complications arise from interactions and group dynamics. -Possible forecasting bias.

Adopted from: Federal Highway Administration. A Guide to Developing Quality Crash Modification Factors. 2010. Federal Highway Administration. Report No. FHWA-SA-10-032. Gross F., Persaud B., Lyon C

Stand-alone multivariate CPMs (1/2)

- These models are conceptually more complex than the SPF-CMF approach, as it is required to incorporate all significant explanatory variables in model, and this is probably the most challenging aspect when developing these models.
- CMFs can be extracted from the model's covariates.

- Negative Binomial and Poisson are the most commonly developed models.
- In terms of statistical modeling, several approaches have been used to handle specific data issues (e.g., many zero instances).
- While these models tend to be accurate and provide a better fit, they require multiple trials for: (a) model format and (b) type of independent variables.

Stand-alone multivariate CPMs (2/2)

Model	Crash type	Statistically significant explanatory variables
P	single-vehicle crashes	daytime, volume/capacity ratio, shoulder width, presence of intersections and driveways, presence of passing lanes
	multi-vehicle crashes	daylight conditions, number of intersections and driveways.
NB	POD crashes	Horizontal alignment, speed limit, visibility, road surface condition, AADT
RENB	POD crashes	
NB	Injury crashes	Horizontal alignment, speed limit, visibility, AADT
RENB	Injury crashes	Speed limit, AADT
NB	all crashes	AADT, lane width, horizontal curvature, vertical curvature, density of pedestrian crossings, density of access points
NB	all crashes	AADT, curve ratio, speed differentials density
NB	all crashes	AADT, average curvature change rate, shoulder width, forest environment
P	all crashes	Season (snow, dry), median width, three traffic lanes, grade
REP		Season (dry, snow), grade, three traffic lanes
Spatial		Season (snow, dry), degree of curvature, median width, three traffic lanes
Spatial	all crashes	AADT, segment length, delay, speed limit
P and NB	all crashes (curve)	AADT, segment length, radius of horizontal curves
	all crashes (tangent)	AADT, segment length, presence of junctions
NB	all crashes	AADT, differential speed, difference in friction, grade, tangent length
NB	all crashes	AADT, segment length, horizontal curvature

Emerging trends: Machine learning-based CPMs (1/2)

- Machine learning (ML) models have been used in the context of crash prediction modeling.
- The great majority of research has focused on classification using a great array of variables (geometric and operational characteristics, weather information, crash characteristics, etc.):
  a. Injury severity classification → fatality/injury/PDO or casualty/no casualty
  b. Real-time crash prediction → crash events / non-crash events

- More limited are the efforts in crash frequency prediction using ML models due to the fact that ML models have greater potential and options for classification.

- SVM regression and XGBoost are the most common frameworks, while Neural Networks that incorporate Poisson/Negative Binomial regression come second.

Emerging trends: Machine learning-based CPMs (2/2)

<u>Strengths</u>

- ML-based methods have a great potential as they are not restricted by insignificant results and maximum likelihood converging issues.
- They are ideal for large datasets, which is the case with datasets that contain real-time information.

<u>Limitations</u>

On the other hand, they are limited as:

- They are a "black box" approach and are not easily interpretable models. To deal with this issue, extensive sensitivity analysis is needed.
- There is no framework to guide the development of ML-based models for crash prediction.
  - *The ACS20 TRB Committee on Safety Performance and Analysis has pointed out this limitations and has plans to initialize the development of such a framework.*

Conclusions

This review summarized more than 100 scientific papers and reports on microscopic crash prediction modeling.

- The HSM stands as the most complete and coherent approach however, it has important transferability, data needed and expertise-related limitations that stop its wider adoption by practitioners.
- Stand-alone crash prediction models consist of several independent variables, that are highly dependent on the available data. It is hard to define the most appropriate model specification while, there are still issues related to the models' transferability.
- Machine learning frameworks are gaining popularity and future research should focus on the development of a framework for those models and suggest ways to improve their interpretability.

RSS 2022
8th Road Safety & Simulation International Conference
Road Safety and Digitalization
08-10 June, 2022 • Athens, Greece

A state-of-practice review on Crash Prediction Modelling

Katerina Deliali, PhD
Research Associate

Together with:
Anastasios Dragomanovits, Ourania Basta & George Yannis

National Technical University of Athens
Department of Transportation Planning and Engineering