Accident prediction in European countries – development of a practical evaluation tool

Francesca La Torre\textsuperscript{a}, *, Niccolò Tanzi\textsuperscript{a}, George Yannis\textsuperscript{b}, Anastasios Dragomanovits\textsuperscript{b}, Thomas Richter\textsuperscript{c}, Stephan Ruhl\textsuperscript{c}, Niovi Karathodorou\textsuperscript{d}, Daniel Graham\textsuperscript{d}

\textsuperscript{a} University of Florence, Via S. Marta, 3, 50139, Firenze, Italy
\textsuperscript{b} National Technical University of Athens, 5, Heroon Polytechniou st., 15773, Athens, Greece
\textsuperscript{c} Technische Universität Berlin, Gustav-Meyer-Allee 25, 13355, Berlin, Germany
\textsuperscript{d} Imperial College London, SW7 2AZ, London, United Kingdom

Abstract

The aim of this paper is to describe the accident prediction procedure and evaluation tool developed within the PRACT project for rural freeways and two lane rural highways. This allows to undertake the following functions:
\begin{itemize}
  \item adapt the base model to local conditions based on historical data;
  \item identify the Crash Modification Factors (CMFs) that could be relevant for the specific application;
  \item verify if the selected CMFs are transferable to the specific condition;
  \item apply the calibrated model to the specific location to be analysed.
\end{itemize}

Different countries, as well as different road authorities within a country, have different levels of expertise and different data availability. The system is therefore structured with different calibration levels ranging from a total lack of historical data to situations where crash data, traffic data and geometric data are all available.

Keywords: Road infrastructure; Safety measures assessment; Accident Prediction Model (APM); Crash Modification Factors (CMF), Transferability

* Corresponding author. Tel.: +39-055-275-8871; fax: +39-055-275-8871.
E-mail address: francesca.latorre@unifi.it
1. Introduction

To improve Road Infrastructures Safety Management, road authorities and road designers need prediction tools allowing them to analyse the potential safety issues, identify safety improvements and estimate the potential effect of these improvements in terms of crash reduction.

An inquiry conducted over 20 different countries has shown that, despite recent advances in the field of accident prediction modelling, most National Road Administrations (NRAs) and other organisations do not systematically use accident prediction methods during decision making for the implementation of road safety treatments. Furthermore, the use of APMs in decision making is more common in countries that have approved guidelines or manuals, which are normally related to a more advanced road safety culture.

Within this framework, the project PRACT (Predicting Road ACcidents - a Transferable methodology across Europe) was funded by the National Road Authorities of Germany, Ireland, UK and Netherlands within the Conference of European Directors of Roads (CEDR) 2013 Transnational Research Programme - Safety. The research partners of the PRACT project are Università degli Studi di Firenze (Project Leader), National Technical University of Athens, Technische Universität Berlin, and Imperial College London. The project aims at developing a European accident prediction model (APM) structure that could be applied to different European road networks with proper calibration (La Torre, Domenichini et al. (2016)).

The core principles behind the PRACT project structure are that:

- it is unrealistic to think that one unique Accident Prediction Model (APM) model with a unique set of Crash Modification Factors (CMFs) can actually be developed, valid for all Europe and for all the different types of networks of motorways and higher ranked rural roads;
- the development of a specific APM model and a set of CMFs based on local data is extremely time consuming and expensive and requires data and experience that most road administrations do not have;
- the development of “local” CMFs only based on historical local data prevents the possibility of evaluating the effectiveness of new technologies.

The basic assumption on which the PRACT project is therefore built is that APMs and CMFs can be transferred to conditions different from the ones for which they have been developed if selected based on scientifically valid criteria and adapted to local condition based on historical crash data.

The PRACT project was aimed at addressing these issues by developing a practical guideline and a user friendly tool that will allow different road administrations to:

- adapt the basic APM function to local conditions based on historical data;
- identify the CMFs that could be relevant for the specific application;
- verify if the selected CMFs are transferable to the specific condition;
- apply the calibrated model to the specific location to be analysed.

This approach acknowledges that different countries, as well as different road authorities within a country, have different levels of expertise and different data availability, and will allow calibration levels ranging from a total lack of historical data to situations where crash data, traffic data and geometric data are all available.

To this aim the PRACT Project followed a logical process in 7 steps as described below:

I. collecting and analysing the APMs currently used by different national road administrations (NRAs) in Europe and worldwide, as well as the currently used data sources for the development and application of APMs. The different APMs have reviewed and assessed in terms of theoretical approaches, characteristics of models in use, implementation conditions, data requirements and available results, with focus on motorways and higher ranked rural roads;

II. proposing the functional structure of the APM to be implemented in the Guideline;

III. reviewing recent and salient literature related to the CMF, including the background and development of the CMF, various methods for developing CMFs, and key issues in the application of the CMF;

IV. identifying key CMFs which have not been fully studied or omitted in the literature and, if possible, developing new CMFs;

V. define the criteria for assessing the transferability of CMFs;

VI. produce a Guideline for the implementation of selected accident prediction models for rural freeways and two lane rural highways and for the evaluation of the transferability of these models to a given road network;

VII. produce a user-friendly tool for calibrating the APM to local conditions and for selecting the CMFs applicable to the specific network.
2. Background

To improve Road Infrastructure Safety Management the road authorities and the road designers need prediction tools allowing them to analyse the potential safety issues, to identify safety improvements and to estimate the potential effect of these improvements in terms of crash reduction. For this aim in 2010 the AASHTO Highway Safety Manual (HSM) was released including a very comprehensive set of models for predicting an estimation of the number of expected road crashes for two-lane rural highways, multilane rural highways, and urban and suburban arterials (AASHTO (2010)).

A first study addressing the issue of the transferability of the rural two-lane two-way roads model to the European networks has been conducted by Martinelli et al. (2009) with reference to the Italian road network of provincial roads.

Crash Prediction Models (usually called also Accident Prediction Models) for freeways were developed by Hasson et al. (Hasson et al. 1998). Persaud and Dzbik (1993) developed two prediction models using data from urban freeways in Ontario, Canada: one for the total number of crashes and one for severe (fatal plus injury) crashes only. These models, together with that proposed by Wang et al. (1998), developed for rural divided highways, with characteristics similar to those of rural freeways and few or no access points, were reviewed and modified by Bonneson et al. (2005) to estimate the predicted numbers of severe crashes per year (i.e. fatal and injury crashes). Recently, Park et al. (2010) have found that the number of predicted crashes is significantly related to average daily traffic, on-ramp density, degree of road curvature, median width and inside shoulder, number of lanes (for urban freeways), and whether the freeway is in an urban or rural area while off-ramp density was not a statistically significant variable. In 2014, a supplement to the 2010 edition of the HSM has been issued with specific models for freeways and interchanges (AASHTO (2014)). The newly developed HSM Freeway model has been applied in Italy by La Torre et al. (2014). A very extensive review of APMs has recently been published in Yannis et al. (2017). Most of the new Accident Prediction Models have identified the following form as the most suitable for allowing the widest transferability:

$$N_p = N_{spf} \cdot (CMF_1 \cdot \ldots \cdot CMF_m) \cdot C$$

where:

- \(N_p\) = predicted average crash frequency for a specific site;
- \(N_{spf}\) = predicted average crash frequency determined for the base conditions of the Safety Performance Function (SPF). This typically is only a function of traffic volumes and segment length;
- \(CMF_1 \ldots CMF_m\) = crash modification factors (that could be also derived from crash modification functions) accounting for specific site conditions (geometric design, traffic control features etc);
- \(C\) = calibration factor to adjust the SPF for local conditions related to the network where the model is to be applied. This accounts for all the factors that lead to safety differences and that are not considered by the safety prediction methodology itself (differences in climate; differences in animal populations that lead to higher frequencies of collision with animals; differences in driver populations and trip purposes; complexity of the geometric layout; driver attitude and behaviour as the rate of compliance with road code rules; vehicle fleet characteristics; crash reporting practices; differences in road standards).

The studies conducted on the Italian network have shown that a single calibration coefficient for the whole prediction model might be insufficient to adapt the HSM models to local conditions that differ considerably from those where the models have been developed.

Crash modification factors and crash modification functions – the indicators that quantify the expected crash variations due to the differences between the base conditions and the specific site conditions (geometric design, traffic control features etc) – are the basis for evidence based safety policies. Specifically, CMFs are fundamental to identifying the most effective road safety countermeasures. Furthermore, they are a useful tool for achieving optimal use of resources as they allow for calculating safety benefits in economic analyses of safety policies. Through a crash modification function (CMF) it is possible to combine different evaluation results and consequently better comprehend and implement effective safety measures (Hasson et al. (2012)). A CMF could allow more rapid adoption and dissemination of new safety measures. The narrower the CMF distribution, the larger is the probability that policy decisions are correct. The US Federal Highway Administration has developed a very comprehensive CMF clearinghouse (www.cmfclearinghouse.org) where CMFs developed worldwide are classified and assessed with a “star rating” approach, but there are several CMFs still missing.

For the prediction of expected crashes in tunnels most APMs available worldwide are not applicable. The most used model is the one developed by Salvisberg et al. (2004) that was developed analysing Swiss roads. The applicability of the Swiss model to the Italian motorway tunnels has been studied and presented in Domenichini.
et. al. (2012). The results show that the Swiss model fits quite well also the Italian existing tunnels even though it is not structured to consider different safety treatments, as those that equip the new tunnels.

A uniform European approach to accident prediction modelling has been recently developed within the SAVeRS Project (Selection of Appropriate Vehicle Restraint Systems) but is limited only to Run Off Road Crashes (La Torre et al. (2015) and La Torre, Erginbas et al. (2016)). Furthermore, the SAVeRS procedure accounts for the calibration of the base Safety Performance Functions and of the overall predicted number of crashes to local condition, but the CMFs are assumed to be transferrable from one country to another.

3. The PRACT Accident Prediction Models

The main outcome of the project was the development of different APMs for motorways and two-lane, two-way roads applicable to different European countries, that constituted the core of the transferring procedure finally proposed in the guidance document and implemented in the computer-based tool. Only fatal-injury accidents were considered. Predictive models consisted of base safety performance functions (SPFs), crash modification factors (CMFs) and a calibration factor (C), according to the formula:

\[ N_{\text{predicted}} = N_{\text{spf}} \cdot \prod_i CMF_i \cdot C \]

Base SPFs were developed for different European countries using generalized linear models with negative-binomial distributions. Different SPFs were calculated using the base datasets (subset of the full dataset that include only segments with specific base conditions where all the CMF values are 1) for single-vehicle motorway accidents, multi-vehicle motorway accidents and rural two-lane, two-way road accidents. A selection of Crash Modification Factors (CMFs) was considered in the model, both for freeways and rural two-lane two-way roads, choosing both valuable CMFs identified in the literature review and CMFs developed internally within the PRACT Project (Karathodorou et al (2015), Karathodorou et al. (2016), La Torre et al. (2017)). A calibration factor was finally calculated for the different European countries and infrastructure types by applying the model to the full dataset. The goodness of fit of both the base models and the final models were evaluated with a Pearson’s chi-square test at a 5% significance level.

In some cases, the size of the base dataset was not sufficient to develop new base SPFs: in these situations, a different base SPF was chosen from the ones developed for different countries and the calibration procedure was limited to the calculation of a calibration factor.

The general structure of the base SPFs was:

\[ N_{\text{spf}} = L \cdot \exp(a + b \cdot \ln(c \cdot AADT)) \]

where \( L \) is the length of the segment [km], \( AADT \) is the average annual daily traffic [veh/day], \( a \) and \( b \) are regression parameters and \( c \) is scaling factor.

Regarding the CMFs, the ones selected for the prediction models for motorways were:

- Horizontal curvature
- Lane width
- Inside shoulder width
- Median barrier
- High volume
- Outside shoulder width
- Shoulder rumble strips
- Outside barrier
- Outside clearance
- High friction wearing course
- Average speed enforcement (section control)

The CMFs selected for rural two-lane two-way predictive models were:

- Road width
- Horizontal Curvature
- Vertical curvature (grade)
- Percentage of heavy goods vehicles
- Shoulder width and type
- Driveway density
- Centerline rumble strips
- Passing lanes
- Two-way left-turn lanes
- Roadside design
• Lighting
• Automated speed enforcement

The different models’ coefficients as well as the calibration factors developed in the PRACT project are shown in Table 1 and Table 2 respectively for freeways and rural two-lane two-way roads.

The Pearson’s chi-square X) values as well as the chi-square-limit values for a significance level of 0.05 (X*(0.05,df)) are also presented. If the Pearson’s chi-square value is lower than the limit value, it’s possible to state that the model has a good fit to the observed accident occurrence.

The models have been improved, with respect to the preliminary models developed within PRACT, by expanding the dataset (UK) or by analysing the dataset in subsets to identify anomalous trends (Italy).

For the base model, all Pearson’s chi-square values were below the 0.05 limit values, so it’s possible to state that the goodness of fit for all of the base models were satisfactory. For the full model, all the models are statistically significant with the only exception of the Netherlands freeway models and the German rural two-lane two-way roads that are not significant with a 95% confidence interval. For these models, an increase of the dataset or a split in more homogeneous sub-datasets should be considered.

It is interesting to observe that the full models developed with full data from one country and the base model from another (e.g. Greece for motorways with the base model based on German data) were still significant. This means that if the calibration dataset is consistent and without anomalies, the use of a European model as a base model can still work properly if sufficient base sections are not available in the specific country.

Table 1. – Freeway model coefficients (a,b,c base model parameters, k – inverse dispersion parameter, C – calibration coefficient, X – Pearson’s chi-square, df – degrees of freedom, X* - chi-square value for a p-value of 0.05 and df degrees of freedom).

<table>
<thead>
<tr>
<th></th>
<th>IT</th>
<th>GE</th>
<th>GR</th>
<th>UK</th>
<th>NE</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>-10.05</td>
<td>-10.47</td>
<td></td>
<td>-2.946</td>
<td>-2.792</td>
</tr>
<tr>
<td>b</td>
<td>1.955</td>
<td>1.523</td>
<td>1.476</td>
<td>1.173</td>
<td>0.158</td>
</tr>
<tr>
<td>c</td>
<td>0.002</td>
<td>0.002</td>
<td>0.002</td>
<td>0.002</td>
<td>0.002</td>
</tr>
<tr>
<td>k</td>
<td>0.861</td>
<td>0.771</td>
<td>4.069</td>
<td>1.318</td>
<td>3.646</td>
</tr>
<tr>
<td>C</td>
<td>1.74</td>
<td>1.175</td>
<td>1.577</td>
<td>0.928</td>
<td>0.464</td>
</tr>
<tr>
<td>X</td>
<td>694</td>
<td>711</td>
<td>1195</td>
<td>1280</td>
<td>108</td>
</tr>
<tr>
<td>df</td>
<td>714</td>
<td>883</td>
<td>1686</td>
<td>1830</td>
<td>91</td>
</tr>
<tr>
<td>(X*(0.05,df))</td>
<td>780</td>
<td>953</td>
<td>1783</td>
<td>1930</td>
<td>114</td>
</tr>
</tbody>
</table>

Base Model calibration dataset description

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Range of traffic (min/max AADT (vehicle/day))</td>
<td>7522/ 49572</td>
<td>6863/ 49572</td>
<td>6704/ 45443</td>
<td>5355/ 45443</td>
<td>-</td>
</tr>
<tr>
<td>Number of segments</td>
<td>138</td>
<td>385</td>
<td>249</td>
<td>906</td>
<td>-</td>
</tr>
<tr>
<td>Number of crashes</td>
<td>94</td>
<td>727</td>
<td>71</td>
<td>823</td>
<td>-</td>
</tr>
</tbody>
</table>

SV = Single vehicle crashes - MV = Multiple vehicle crashes
Table 2. – Rural two-lane, two-way roads full model coefficients (a,b,c base model parameters, k – inverse dispersion parameter, C – calibration coefficient, X – Pearson’s chi-square, df – degrees of freedom, X* - chi-square value for a p-value of 0.05 and df degrees of freedom).

<table>
<thead>
<tr>
<th></th>
<th>IT</th>
<th>GE</th>
<th>UK</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>-7.363</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>0.805</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>k</td>
<td>0.307</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>0.397</td>
<td>1.064</td>
<td>0.559</td>
</tr>
<tr>
<td>X</td>
<td>602</td>
<td>12284</td>
<td>106</td>
</tr>
<tr>
<td>df</td>
<td>753</td>
<td>8763</td>
<td>105</td>
</tr>
<tr>
<td>(X*(0.05,df))</td>
<td>818</td>
<td>8982</td>
<td>135</td>
</tr>
</tbody>
</table>

Base Model calibration dataset description

| Years of observation | - | 2010-2014 | - |
| Range of traffic (min/max AADT [vehic/day]) | - | 507/17040 | - |
| Number of segments | - | 458 | - |
| Number of crashes | - | 269 | - |

4. CMF transferability

Within the PRACT project a procedure to assist a practitioner to implement a specific treatment on a road segment in his local area where no CMF based on the local conditions is available, has been developed. Depending on the CMFs present in the literature for the specific treatment, the data availability for the local road network where the practitioner wants to implement the treatment and the comparison of the local conditions with the road characteristics of the road segments for which the CMFs have been developed, the practitioner can face with one of the following situations:

Situation a - A CMF estimate θ is available in the literature from another jurisdiction. Some descriptive statistics for key characteristics of the road network used in its estimation are also available.

**Solution** - The practitioner should use the available estimate θ if the values of key characteristics of the road segment in his local area fall within the range of values used in θ’s estimation (i.e. between the minimum and maximum value). Ideally, the value should fall within the 5th and 95th percentile of the values of the road sample used in CMF estimation, however this information is generally not readily available in studies.

In the PRACT tool this comparison is made automatically, also identifying the deviation from the CMF development range. If this limit is within 10% of the limiting value a warning is given in yellow. For larger deviances the warning is given in red.

Some indicators are given in a qualitative form as they can be present only in some of the segments used for the CMF development. These are defined based on the frequency of occurrence as Always/Frequently/Rarely/Never occurring. If the segment has a given feature (e.g. Rumble Strips) the CMF is considered consistent if it was developed on a network where these occur Frequently or Always. If the segment does not have the feature the CMF is considered consistent if it was developed on a network where these occur Rarely or Never.

Situation b - A CMF estimate θ is available in the literature that has been estimated on conditions different from local conditions (or for unknown conditions). The practitioner also knows that the treatment has already been
implemented in \( n \) road segments in his local area. There are available data on accident rates on these segments before the treatment \((X_1, X_2, \ldots, X_n)\) and after the treatment was implemented \((X'_1, X'_2, \ldots, X'_n)\). The practitioner wants to determine whether \( \theta \) is applicable to their local conditions.

**Solution** - If \( \theta \) is valid in the practitioner’s local conditions, then the expected number of accidents after treatment implementation should be \( \theta X_1, \theta X_2, \ldots, \theta X_n \). The idea is to compare these expected accident rates with the observed accident rates \( X_1', X_2', \ldots, X_n' \). This can be done using a paired \( t \)-test as follows:

Calculate the difference \( d_i = X'_i - \theta X_i \)

Calculate the \( t \)-statistic as \( T = \frac{\bar{d}}{s_d / \sqrt{n}} \)

where \( s_d \) is the standard deviation of \( d \).

If there is no difference between \( X'_i \) and \( \theta X_i \), then \( T \) follows a \( t \)-distribution with \( n - 1 \) degrees of freedom. If \( T \) is significant then \( \theta \) is not applicable in the practitioner’s local area. If \( T \) is not significant, then \( \theta \) is applicable in the practitioner’s local area.

If the above analysis suggests that \( \theta \) is not applicable in the practitioner’s local area, then we suggest the use of a composite CMF estimate based on CMF estimates from several countries/jurisdictions using the formula provided in ‘situation c’ or use of a naïve before-after estimate based on local accident data. The applicability of the composite CMF estimate can also be checked using the analysis above.

**Note:** The methodology is applicable for treatments where accident rates before and after the treatment are available. For road characteristics where cross-sectional data are available for different values of the road characteristic (e.g. horizontal curvature) the methodology is not applicable. Moreover, a similar methodology cannot be developed due to the fact that road segments in the sample of local data are likely to have many different characteristics in addition to the road characteristic under consideration. Practitioners should proceed as in ‘situation a’.

**Situation c** - A practitioner wants to implement a specific treatment on a road segment in the local area, and although \( n \) CMFs are available in the literature \((CMF_1, CMF_2, \ldots, CMF_n)\) no information or local data are available to test whether these are applicable to the practitioner’s local area.

**Solution** - A composite CMF estimate, \( CMF_{comb} \), can be estimated as a weighted sum of available CMFs. The corresponding formula is:

\[
CMF_{comb} = \sum_{i=1}^{n} \frac{1}{S_i^2} \cdot CMF_i
\]

Where \( CMF_i \) is the \( i \)th CMF estimate and \( S_i \) the standard error of the \( S_i \) estimate.

The standard deviation of the combined CMF can be estimated using the formula below:

\[
\text{Var}[CMF_{comb}] = \sum_{i=1}^{n} \frac{1}{\sum_{i=1}^{n} \left( \frac{1}{S_i^2} \right)}
\]

In the PRACT tool this combination is made automatically if the user wants to combine CMFs that are already in the tool dataset. Additional CMFs can be combined by adding them into the dataset prior to applying the CMF combination part of the tool. Note that only CMFs for which the standard deviation is given can be combined.

5. The PRACT Tool

The APMs and CMFs were implemented into the computer-based tool and the user is only required to describe the characteristics of the road section in order to correctly estimate its predicted average crash frequency.

The user can choose one of the base models and calibration coefficients already included in the tool (the coefficients of which are given in Tables 1 and 2 above), but can also add a new base model or add a new calibration coefficient based on the calibration on the local network of one of the PRACT base models. This allows the system to be extremely flexible and easily updatable.

Once the user chooses the CMFs to apply the tool automatically controls if the situation that the user is analysing belongs to the range of applicability of the given CMF, otherwise a warning message is given. Additional notes are given (e.g. if the CMF applies only to a subset of crashes).
As mentioned earlier, one of the key features in the tool is the implementation of the transferability checks (situation “a”) and the meta-analysis for combining different CMFs available for the same feature.

Figure 1 shows a screenshot of the control checks performed by the tool in a specific application for one of the Section Speed Control CMFs available in the tool.

As far as for this specific feature more than one CMF is available (with both the CMF value and the standard deviation), for this feature the user can decide to use a combination of the different CMFs obtained with the two with a meta-analysis approach (situation “c”) as shown in Figure 2.
The final version of the PRACT Models has been recently approved by the Project Manager and the final guideline and tool will be available online (www.practproject.eu) by the end of 2017.

6. Conclusions

The extensive background evaluation conducted in the PRACT project has shown that there is a strong need for a practical and harmonized approach to accident modelling. The most suitable functional form to obtain a transferrable accident prediction model includes a base safety performance function and a set of CMFs.

A set of base models for freeways and rural two-lane two-ways roads have been produced, all statistically significant. These have been then calibrated with larger datasets with “non base” conditions to have a set of full models that can be applied to any section of the analysed networks.

The procedure developed can be applied at different levels based on the data availability and expertise and has shown that using a European base model calibrated with a proper local dataset from another country can lead to statistically significant full models.

To allow the user to identify the proper CMFs to be applied and to combine the existing ones, if needed, a CMF transferability procedure has been defined and implemented in the PRACT tool.

7. Acknowledgments

This research was carried out within the project PRACT - Predicting Road Accidents - a Transferable methodology across Europe funded by the National Road Authorities of Germany, Ireland, UK and the Netherlands within the Conference of European Directors of Roads (CEDR) 2013 Transnational Research Programme - Safety. The authors wish to thank the Project Manager (Elizabeth Mathie) for her support and advice.

8. References


Karathodorou, N., Graham, D., Hu, J., Richter, T., Ruhl, S., Yannis, G., Dragomanovits, A., Laiou, A., La Torre,


