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

A state-of-the-art Review on Crash Occurrence Analysis and Hazardous Location Identification

Paola Tiberi[1], Antonino Tripodi[1], Anastasios Dragomanovits[2], Aikaterini Deliali[2], George Yannis[2]

[1]FRED Engineering, 15 Via Celimontana, 00184, Rome, Italy
[2]Department of Transportation Planning and Engineering, National Technical University of Athens, 5, Heroon Polytechniou st., 15773, Athens, Greece

Abstract

Road crashes are a leading cause of deaths for certain age groups, road user types, and countries, while crash-related injuries have been associated with an array of negative impacts for both the affected individuals and the society. The identification of hazardous road sections stands as one of the most applied processes in successful road safety management and effective methods to correctly identify those locations are essential for road agencies and safety experts. This study provided a state-of-the-art review of international literature on methods and applied practices related to the identification of hazardous road sections. Emphasis was also given to the applied practices across the EU countries which in some cases was feasible through non-English manuals and guidelines; many resources were identified through a questionnaire survey targeted to road safety practitioners and relevant stakeholders across the EU countries. The analysis identified and summarized differences and similarities in the various methods and applied practices and therefore, the reviewed studies are classified in groups with respect to (a) the used safety performance metric (e.g., crash rates vs crash frequency), (b) criterion to prioritize road sections/sites for remedial treatments (e.g., sites that exceed the area-wide crash rate will be prioritized), (c) type of crash data (e.g., fatal crashes or fatal and severe injury crashes), and (d) area of application which refers to the part of the road network such as road segments and intersections. The findings indicate that the analysis usually relies on 3 to 5 years of crash data with different levels of detail; however, there are cases that use less than 3 years of crash data. Moreover, the identification of hazardous locations is a widely applied method in road safety management, however, even within the same country different road authorities choose different safety performance metrics to identify hazardous locations or prioritize locations for road safety improvement using different criteria. This work is useful for practitioners so that they can identify potential ways of improving their applied crash occurrence methodology, while it will inform researchers on paths for future research in the field of crash analysis.

Keywords: hotspot analysis; blackspot analysis; hazardous road locations identification; hazardous road sections identification.

---

*Aikaterini Deliali Tel.: +30-6971551576;
E-mail address: kdeliali@mail.ntua.gr

1. Introduction

Road authorities and traffic engineers employ crash occurrence analysis with the objective to identify crash contributing factors and in turn select appropriate countermeasures. The analysis of crash records can reveal crash contributing factors related to road user behavior (e.g., seat belt or helmet use), vehicle type and vehicle conditions (e.g., motorcyclists or old vehicles are more prone to injury-related crashes), external conditions (e.g., weather and lighting conditions, time of the day, etc.), and the road infrastructure. For the assessment of road infrastructure, crash occurrence analysis is implemented to (1) identify high-risk or hazardous or unsafe locations across a network and (2) to evaluate the effectiveness of countermeasures.

Methodologies to identify hazardous locations may be split in two categories depending on their scope. A methodology may seek to identify unsafe/hazardous sites or in other words, locations of a relatively small size, such as an intersection. This process is also known as “black spot identification” or “hot spot identification”. In this review this will be noted as site-level analysis. On the other hand, a methodology may examine and in turn seek to characterize as safe or unsafe longer sections, usually more than one kilometer; this will be noted as network-level analysis.

Regardless the focus of the analysis, the identification of hazardous locations is a critical step of a road safety management process as it indicates which sections should be subject to safety improvements to reduce or even eliminate crashes between two or more drivers or drivers and fixed objects. The correct identification of these sections is of high importance. Identifying truly hazardous locations is the first step in eliminating road injuries and casualties. Falsely identifying a location as hazardous (while it is not) will likely result in ineffective allocation of funding for road safety improvements. Across the globe, multiple methodologies have been developed and are currently implemented for the identification of hazardous locations using crash records analysis. These methodologies have a similar structure, consisting of four stages. Essentially the differences between various methodologies can be found in the completion of each stage. First, the network of interest is broken down in smaller parts a process known as segmentation. Then the second stage concerns the estimation of the safety level of each part of the network based on a crash-based metric (or safety performance metric). The third stage of the hazardous location identification process is the definition of a criterion (or more) based on which a section is classified as hazardous.

This review presents existing methodologies for the identification of hazardous locations across motorways and rural roads. The methodologies have been identified through national guidelines, project reports, scientific journal papers and the objective was to cover the international literature however, more focus was given to the European countries. A questionnaire survey was disseminated to European road authorities, research institutes, academic researchers, and road safety experts to ensure that resources not available in English language and so, not easily identified through online search engines, will also be considered.

The rest of the paper is organized as follows: Section 2 gives an overview of the questionnaire survey while Section 3 presents an overview of currently applied methodologies on crash occurrence analysis across the European counties. In Section 4 the main steps for the identification of hazardous locations are presented and discussed. Section 5 discusses available methodologies and provides points for consideration. The contributions of this work as well as future extensions are presented in the Conclusions (Section 6).

2. Questionnaire survey

A questionnaire survey was designed with the objective to collect information on currently applied road safety practices and methodologies across Europe. As many European countries follow their national methodology, which is written in the country’s official language, it is not always feasible to get access to these guidelines through online search engines. The questionnaire survey consisted of four parts: (1) types of roads assessed with the national methodology, (2) available road types and storing systems used for the analysis of road safety analysis, (3) applied methodologies and practices for the assessment on roads using crash data, (4) applied methodologies and practices for the in-built safety assessment of roads. In this paper, the focus is mainly on the 2nd part of the questionnaire survey that concerns the crash-based methodologies.

A total of 35 responses were received representing the following 29 countries: Austria, Belgium, Bulgaria, Croatia, Cyprus, Czechia, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Norway, Poland, Portugal, Slovenia, Slovakia, Spain, Sweden, and United

Kingdom (UK). For some countries, more than one response was received, representing different stakeholders (e.g., a road authority and a research institute). Overall, many of the responses were from public authority representatives, i.e., Ministry or National Road Authority (NRA), as shown in Figure 1.

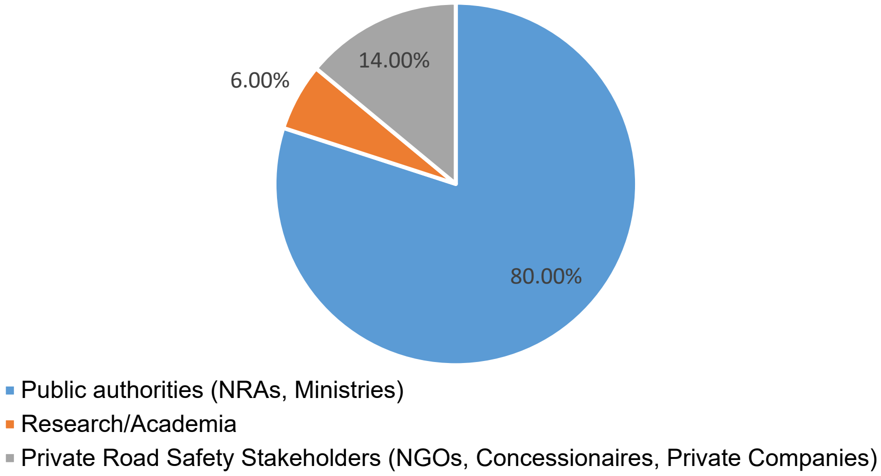

Figure 1. Questionnaire respondents by organization type.

3. Crash occurrence analysis methodologies

Through the questionnaire survey, most countries provided an overview of their approach to assess road safety, while some countries also provided their national guidelines. Additional methodologies were identified through an online search. An overview of the identified methodologies is presented in Table 1.

Table 1: Overview of applied crash occurrence methodologies

Country	Method	Network screening/ segmentation	Criterion	Time period
Austria	Crash density	Sliding window (250m)	3 Injury Crashes 5 All Crashes*	3 years 1 year
Germany	Crash frequency	Small sections (no defined length)	5 All Crashes 5 Injury Crashes 3 Serious Injury Crashes	1 year 3 years 3 years
Malta	Crash frequency	Rural junction Rural segments	5 Injury Crashes 3 All Crashes	5 years 1 year
Hungary	Crash density	Sliding window (1.000m)	4 Injury Crashes	3 years
Denmark	Crash density/rate	Sliding window (400-500m)	5 All Crashes	3 to 5 years
Czech Republic	Crash frequency	Intersections and road sections (<250m)	3 Injury Crashes 3 Injury Crashes 5 All Crashes	1 year 3 years 1 year
Croatia	Crash rate	Sliding window (300m)	1 Fatal Crash Critical level	3 years 3 years
France	Crash frequency /density	-	5 All Crashes Density of the section	5 years
Lithuania	Crash density/rate	Sliding window (500 m)	3 All Crashes Critical level	4 years
Belgium Flanders	Priority value	Sliding window (100m)	3 Injury Crashes	3 years
Spain	Crash density/rate	3 km long sections	5 Injury Crashes 10 Injury Crashes Danger Index	3 years

	Potential reduction in crash costs	homogenous sections (>5km)	Sections are sorted based on their crash reduction potential	3 years
Slovakia	Crash frequency	*Not reported*	Percentage of fatal crashes	1 to 3 years
Slovenia	Crash frequency/rate	*Not reported*	*Not reported*	3 years
Italy	Crash rate Crash frequency Crash number	Homogenous sections	Sorting the values of the resulting indicators (primarily fatal crashes)	3 years
Bulgaria	Crash rate	300m-1.000m long sections	Predefined thresholds for crash rates	1 year
Poland	Crash rate	2-10km long sections	Individual risk classes (based on fatal crashes)	3 years
	Standardized crash cost density index	2-10km long sections	Definition of risk classes based on crash cost rates (based on all crash types)	3 years
Denmark	Crash density/rate	5km sections	Average crash density/rate	3 to 5 years
France	Crash density/rate	3-20km long homogeneous sections	Average crash rate (at the national level of same road types)	5 years
Ireland	Crash frequency/rate	1km long sections	Average crash frequency and rate for the Reference Population	3 years
Iceland	Crash frequency/rate	*Not reported*	Average crash frequency and rate for the particular road type	5 years
Germany-France	Safety potential	Section > 3km long sections with at least 3 serious injury crashes	Sorting the sections based on the safety potential reduction	3 to 5 years
Austria	Crash cost rate	3km long sections	Crash cost rate classes (based on all crash types)	3 years
iRAP Risk Mapping	Crash density/rate Crash cost density/rate	Homogeneous sections large enough to have 20 fatal or serious injury crashes over the last 3 years	Categorization in five classes	3 years

[*] *All crashes = crashes of types in terms of injury severity level.*

It is evident that while European countries implement crash occurrence analysis for the safety assessment of their roads, there are differences in the applied methodologies. These differences can be seen in terms of performance metrics for the safety assessment (e.g., use of crash rates vs crash density) as well as implemented criteria to classify a section/segment as safe or unsafe (defined thresholds based on single values vs ranges of values), in the years of crash data used for the assessment (ranging from one to five years), and lastly in the implemented segmentation process (sort segments of several hundred meters vs long sections of several kilometers). The following section discusses in more detail the aforementioned differences.

Several countries have reported to use more advanced methodologies for the identification of hazardous locations. Nordic countries like Denmark, Norway, Finland and Estonia rely on locally developed crash prediction models for assessing road safety. Through the questionnaire responses it was not always clear though whether crash prediction models are used proactively or whether the predicted crash frequency is compared to the observed one. In the first case, recent crash data is ignored, and the models indicated that sections with certain characteristics (e.g., traffic volumes, presence of curves or certain speed limits) are less safe as they yield to higher predicted crash frequency compared to the other sections.

It is noted that through the questionnaire, several countries indicated that they are solely rely on non-crash based approaches to safety or they are partially implementing non-crash based approaches as a means of identifying hazardous sections across their network; these will not be discussed in this study but are mentioned for the sake of completeness. Sweden is the only country in Europe that fully relies on proactive safety assessment. In Czechia, a method has been developed to proactively identifying unsafe horizontal curves. Various countries use or have used iRAP Star Rating Protocol for assessing road safety, however, this methodology is used for parts of the network rather than throughout the country.

4. Steps for the identification of hazardous locations

4.1 Network segmentation

The network of interest is divided in smaller sections so that each that safety level of each can be assessed. In the literature it is highlighted that different road facilities, i.e., road segments, intersections or interchanges should be assessed separately [1], as they do not have comparable design and operation characteristics. Additionally, facilities of the same type (e.g., road segments) should be further grouped based on the road type, e.g., differentiate between road segments of a motorway vs a rural road and design, e.g., two-lane vs multi-lane rural roads and four-leg intersections vs roundabouts.

Multiple approaches exist for dividing road sections in smaller parts however, they can be broadly split between two categories: (1) fixed-length sections or (2) homogeneous sections. Diving the network in same length sections is a straightforward and simple process. For site-level analysis, section lengths range from 100m to 1000m while for network-level analysis lengths are above one kilometer. Homogeneous sections are defined based on several variables, most common of which are: traffic volume, number of lanes, lane widths, roadside hazards, and presence of curves. Fixed-length segmentation is a simple approach, while determining homogeneous road sections is significantly more complex even when this depends on a single variable.

4.2 Network screening

In addition, the concept of network screening is essential for the site-level analysis. Screening methods apply to segments and ramps and not intersections that are treated as points rather than continuous elements. The objective of screening is to identify the exact location within the road section (segment or ramp) that is more likely to be benefited by safety improvements. Essentially, the network is broken down into sections and then each section can be further broken down to narrow down the set of spots to be treated. In other words, dividing each section into smaller parts allows to understand what portion of the road section controls the section's critical crash frequency will make it easier and more efficient to identify effective countermeasures, although such a detailed analysis can be time consuming and data intensive.

The network screening methods identified in the review with regards to road sections are the sliding window method, the peak searching method, and the simple ranking method. The node screening method is used for intersections and finally the facility screening is selected when the objective is to screen segments and intersections at the same time.

4.2.1 Sliding window method

In the sliding window method, a window of a specified length moves along the section in specified increments. The selected performance measure is calculated for every position of the window and these results are recorded. Given that the section length does not necessarily fit a certain number of windows, a window can bridge two or more sections that are continuous. A window pertains to a given segment if at least some portion of the window is within the boundaries of the segment. From all the windows that pertain to a given segment, the window that shows the highest safety potential out of the whole segment is identified and is used to represent the safety potential of the whole segment. After all segments are ranked according to the respective highest sub-segment value, those segments with the greatest potential for reduction in crash frequency or severity are studied in detail to identify potential countermeasures.

The benefits of the sliding window method are that it is a more targeted technique for the identification of the hazardous location across a section. On the other hand, it is a time-consuming method, and it requires many calculations and good mapping tools for the network.

4.2.2. Peak searching method

The peak searching method is similar to the sliding window method in the sense that the road section is divided into smaller parts; the HSM advises that the window is equal to 0,1mile. The windows do not overlap with the exemption of the two last ones at the end of the section. This overlapping is intendent to occur within the section so that the last window of the first section does not overlap with the first window of the following section. The window with the highest value (i.e., safety performance metric) is chosen to represent the entire section. Before concluding on the final safety performance metric value for the section, a test needs to be conducted. It is known as “precision test” and is given by the following formula (1):

$$CV = \frac{\sqrt{Var(Performance\ metric)}}{Performance\ metric} \quad (1)$$

Where CV stands for coefficient variation. A large CV indicates a low level of precision in the estimate, and a small CV indicates a high level of precision in the estimate. The calculated CV is compared to a specified limiting CV. If the calculated CV is less than or equal to the CV limiting value, the performance measure meets the desired precision level, and the performance measure for a given window can potentially be considered for use in ranking the segment. If the calculated CV is greater than the CV limiting value, the window is automatically removed from further consideration in potentially ranking the segment based upon the value of the performance measure. At this point a new window length needs to be set and the process should be repeated based on it.

Szénási et al. compared the different techniques used for screening and the hazardous locations’ identification [2]. The authors conclude that the Peak searching method is one of the simplest, one-dimensional techniques.

4.2.3. Simple ranking method

The simple ranking method can be applied to nodes and segments (and ramps). In this method, the performance measures are calculated for all the sites under consideration and in the case of segments and ramps one calculation is done per case and represented the entire length of the road element. The results are then ordered from high to low.

The simplicity of this method is the greatest strength. However, for segments, the results are not as reliable as the other segment screening methods, especially when the length is large.

4.3 Safety performance metrics

Combining various sources consisting of manuals and project reports, the following safety performance metrics have been identified (Table 2).

Table 2: Safety Performance Metrics

Metric	Explanation
[*]Crash frequency	Number of crashes per year
[*]Average crash frequency	Average number of crashes per severela years
[*]Crash density	Average number of crashes per year over the section length; for intersections this is the same as the average crash frequency
[*]Crash rate	Average number of crashes per year over a traffic volume metric usually AADT however it can also be ADT or peak hour volume.
Crash cost	All crashes are translated to monetary terms based on their severity level and finally, the total cost of those crashes is estimated. The crash cost information can also be combined with section length and/or traffic volume information to estimate crash cost density or crash cost rate.
Equivalent Property Damage Only (EPDO)	All crash severity levels (e.g., fatal crashes) are translated into PDO crashes considering the monetary value of each crash type. The nominator of this ratio is the total number of PDO crashes that have occurred, and the denominator is the studied time period. The EPDO information can also be combined with section length and/or traffic volume information to estimate EPDO crash density or EPDO crash rate.
Relative Severity Index (RSI)	RSI combines crash type information (e.g., rear-end crash) with crash severity information, by considering a monetary cost per crash type and estimating the overall crash cost per site.

[*] *When estimating these metrics crashes of all severity levels may be used or the analysis may focus on crashes of certain injury severity levels.*

Crash frequency and average crash frequency are the simplest safety performance metric as they require a minimum amount of data. As crashes are rare and random events it is advised to work with crash data of several years (i.e., at least three years) to eliminate effects like regression-to-the-mean and therefore average crash frequency is used instead of crash frequency. However, some agencies conduct their analysis using the number of crashes recorded in the previous years (e.g., Wallonia region in Belgium). Another note on these metrics is that they are appropriate for junctions and not segments/sections. Average crash frequency is appropriate for junctions as these are road facilities with no linear dimension. In facilities with linear dimension, such as segments and sections, crash density a more accurate metric in detecting segments with high crash frequency, as long segments are expected to have more crashes (per length). Crash frequency can be used instead of crash density when the segmentation method considers segments of the same length.

Crash rate incorporates exposure information, which is a metric of traffic demand. In many cases, and that was evident in the questionnaire responses, agencies refer to crash density as crash rate and essentially consider length as a metric of exposure. In this review, crash rates are metrics that consider some form of traffic volume as exposure. The following equation describes the generic form of crash rates [1]:

$$Crash\ rate = \frac{Average\ crash\ frequency\ in\ a\ period}{Exposure\ metric\ for\ the\ same\ period} \qquad (2)$$

Traffic volume information, as exposure, assuming that roads with higher traffic volume are more likely to experience crashes compared to low-volume roads. Therefore, the same crash frequency results in higher crash rates in a low-volume road compared to a high-volume one. Traffic volume information is usually in the form of Annual Average Daily Traffic (AADT) or Average Daily Traffic (ADT) or peak hour volume. For segments of different lengths crash rates should be based on number of crashes per year over traffic volume and segment length information, i.e., vehicle-kilometer travelled, or vehicle-miles travelled. For intersections traffic volume information is usually in the form of total AADT across all intersecting road or as total entering vehicles (TEV).

While it is relatively straightforward how to account for exposure in the case of motorized vehicle crashes, it is more complicated for the case of crashes between motorized vehicles and pedestrians or bicyclists. These road users are not expected to be present in motorways however, they be present in rural roads and therefore, their safety should be assessed too. In the literature there are three ways for incorporating exposure when studying the aforementioned crashes: (a) motorized vehicle demand, (b) pedestrian or bicycle demand and (c) both motorized vehicle and pedestrian/bicycle demand. Accurate pedestrian or bicycle demand data is hard to obtain and so, it is common to either omit such a term or use road environment factors as proxies of pedestrian/bicycle demand. For example, instead of pedestrian demand data researchers have used the density of crosswalks [3] or population metrics [4]. Fournier et al. showed that for bicycle-motorized vehicle crashes omitting motorized vehicle demand or bicycle demand when estimating crash rates results in over- or underrepresented crash rates, therefore both exposure types are needed [5].

Crash frequency, density and rates may be estimated by using different types of crashes in terms of injury severity. Injury severity information is included in crash occurrence analysis as in many cases the objective is to eliminate injury-related crashes instead of those that result in damages. There are several systems to report injury severity and generally, there are considerable differences between the applied practices across different countries.

The Equivalent Property Damage Only (EPDO) and Crash cost are metrics that produce a final score that represents both crash frequency and severity. Both metrics consider injury severity information and weighting factors relevant to injury severity. EPDO is defined in the HSM [1] and is given by the following equation:

$$Score_{EPDO,i} = f_k N_{k,i} + f_{inj} N_{inj,i} + f_{pdo} N_{pdo,i} \qquad (3)$$

Where f_k, f_{inj}, f_{pdo} are the weighting factors for fatal, injury, and PDO crashes respectively. $N_{k,i}$, $N_{inj,i}$, $N_{pdo,i}$ are the observed fatal, injury, and PDO crashes for the site *i*. The weighting factor for the *j* severity level is estimated as:

$$f_j = \frac{Cost_j}{Cost_{pdo}} \qquad (4)$$

Where $Cost_j$ is the monetary cost of a fatal or injury crash, while $Cost_{pdo}$ is the monetary cost of a PDO crash. Essentially, the weighting factor for PDO crashes is equal to one, while weighting factors for fatal and injury crashes are higher than one. Crash costs are estimated per country or jurisdiction, and generally, there are considerable differences per country. The recent review by Wijnen et al. focused on 31 European countries and found that cost per fatality varies greatly, i.e., between €0.7 million to €3.0 million, while the costs related to serious injury crashes were found to be equal to 2.5% up to 34.0% of the costs of a fatality [6].

4.4 Safety ranking

This subsection focuses on the existing criteria for determining the safety level of a site or a section. First it is noted that according to Table 1, existing methodologies may work with a binary classification, i.e., classifying sections as safe or unsafe based on a single criterion, or there is also the case of multiple safety levels.

Methodologies that work with thresholds are more likely to have a binary safety classification. The thresholds can either be an average value across multiple sites (e.g., national average crash rate across all roads of the same type), or a predefined value (e.g., 3 fatal crashes per section during the past 3 years) determined as critical by safety practitioners (e.g., see the case of Croatia, Hungary, etc. in Table 1). The national average is beneficial as it stands as a straightforward criterion for classifying a section as safe or unsafe. Methodologies that have numerical thresholds (e.g., 3 fatal crashes per section during the past 3 years) do not fully support how this number was determined and so, such thresholds are not easily transferable. Additionally, it is not clear when such thresholds should be revaluated.

It seems that methodologies that work with crash cost rates are more likely to have multiple classes for safety ranking. This is the case for Austria and Poland. In the case of Austria for example, a network-wide classification takes place using six different classes based on crash cost rates [7] (see Figure 2). Non-binary approaches are beneficial in the sense that a better idea is provided regarding the safety level of an entire network and so, better safety management plans can be designed.

5. Discussion

The review reveals that there are numerous methods to assess road safety according to crash occurrence and vary from country to country and even within a country, although they often have a common structure. Some of them cover different aspects of road safety analysis and consist of a multi-step process in which the procedures for identifying high-risk sites, hazardous sections based on crash density, crash rate or other indicators are explained. Identification of high-risk sites has long tradition in traffic safety engineering, but it is evident that different countries do not use fully similar principles. The cause for this can be found in different traffic and technology factors of the countries and in the different levels of their traffic safety.

For all the above-mentioned methods, the availability and reliability of crash data is a significant requirement for identifying hazardous locations or road sections. Moreover, the use of crash data implies the need for statistical tests to make sure that the random variation is not the decisive factor in the process of identifying the high crash frequency locations. However, only a few of the methods (Croatian, French and Spanish) analyzed consider this. Another important aspect is the under-reporting of road crashes that can put into question the results of even methodologically robust approaches.

Since a high-risk site is a site that has an abnormally high number of crashes, some reference to the normal level of safety, i.e., a threshold value, is needed. References to the normal level of safety are generally made by comparing the number of crashes at sites identified as high risk to the number of crashes expected for similar sites, estimated by means of crash prediction models or by referring to a set of normal crash rates. Mainly the identification is based on the number of crashes occurred in a location or a small stretch (sliding window) within a time period. The sliding window method is widely implemented for finding segments that meet the criteria of crash's threshold. In Ghadi and Török (2017), the threshold value is calculated based on crashes frequencies as the average observed number of crashes for all tabulated similar locations with a level of confidence interval [8]. However, in the methodologies analyzed, these threshold values are generally fixed and pre-established.

From the reviewed methodologies, most do not consider the crash severity and when it is considered, there is no standard way of doing so. Three different approaches can be identified, with regards to severity consideration. One approach is to set a more stringent critical value for the number of serious injury crashes than for all injury crashes. A second approach is applying weights to crashes at different levels of severity. A third approach is to estimate the costs of crashes. These costs vary according to injury severity; hence, costs will be higher at sites that have a high proportion of fatal or serious injury crashes.

Considering only the crash density associated with a road section could lead to misleading conclusions when traffic volumes change significantly from one section to another. It is obvious that two different road segments containing the same number of crashes, or the same crash density cannot be equally dangerous when the traffic volumes assigned to each segment vary to a considerable degree. This rational gap is filled by the calculation of the crash rate.

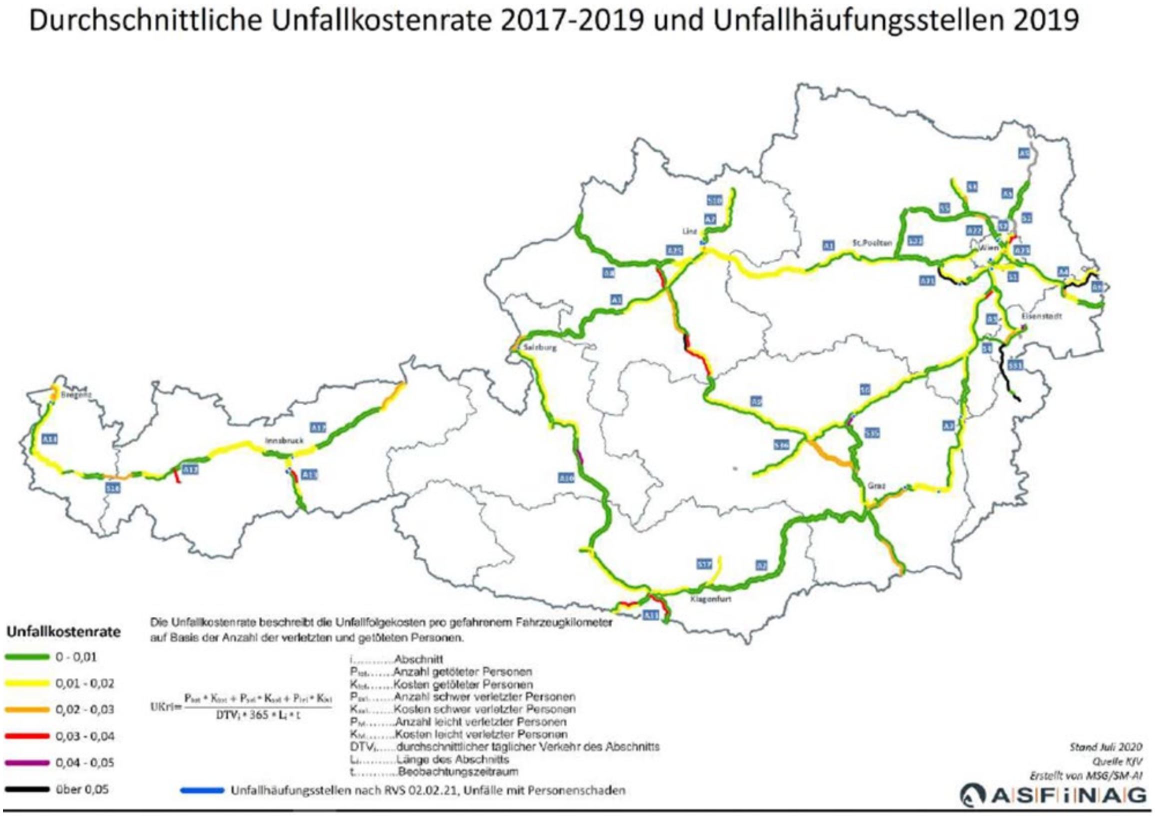

Figure 2. Network-wide safety ranking in Austria's motorways based on crash cost rates [7]

Some methodologies refer to a less punctual analysis, which considers sections and routes of the road network, instead of specific sites. The identification of such network elements is based on crash density, crash rate or other safety-related indicators. Among these, the most frequent is the Safety Potential (SAPO). Brannolte et al. claimed that SAPO is the most important parameter to identify black spots/sections on which safety improvement measures are expected to have the greatest effects [9]. Furthermore, they state that the advantage of the SAPO compared to the classic crash parameters is that it allows assessing different road types and roads with different volumes at the same time. Moreover, as the SAPO is given in crash cost, it can be related to the cost of the improvement measures.

In the African Development Bank Group report on the reactive approaches [10], it is stated that route/corridor analysis is particularly useful since it does not necessitate the precise crash coordinates necessary for blackspot analysis. Route/corridor analysis should be undertaken alongside blackspot analysis since the two approaches will highlight different issues; route/corridor analysis may uncover issues that pertain to longer sections but are not concentrated enough to appear as blackspots. Some information about crash locations is still necessary to attribute crashes to road sections e.g., crash coordinates or road number, road section, link node location, or chainage along a road.

For almost all methods analyzing stretches in this study, the first step is to divide the network into homogeneous sections. Although ideally these road sections should have similar design features and similar traffic flows, should be between 10 km and 150 km in length and as similar in length as possible and meaningful e.g., road between two junctions or between two settlements [11], each country applies different criteria for network segmentation. In particular, in four methodologies, the maximum length of the homogeneous sections is less than 10km, in two only the minimum length is specified, and the remaining methodologies analyze sections that are either longer or of undefined dimension. In general, a variable length of the sections is recommended as long as homogeneous characteristics along them are ensured.

A further aspect to be considered for such safety analyses is the reference period of the crash data. The length of the period used to identify both high risk sites and hazardous segments varies from county to country. The methodologies analyzed cover a period of between one and five years, but most use a three-year database. Nguyen et al. [12] write that research by Cheng and Washington [13] shows that the gain in the accuracy of black spot identification obtained by using a longer period of three years is marginal and declines rapidly as the length of the period is increased. There is little point in using a longer period than five years. Additionally, Land Transport New Zealand stressed out that a three-year crash period could be used in heavily trafficked networks or areas where road changes are recent or ongoing [14]. A three-to-five-year period is preferred for the data because:

- It is long enough to provide a sufficient number of crashes for meaningful results.
- It is short enough to limit the number of traffic and environmental changes that may bias results.
- It helps remove statistical fluctuation and reduce the impact of the regression-to-the-mean effect.
- It provides a consistent base for before and after comparisons.

6. Conclusions

This study reviewed existing methodologies on crash occurrence analysis with the objective to determine the safety level of road sections. Emphasis was given to methodologies applied in European countries and many resources were identified through a questionnaire survey that was disseminated to road safety stakeholders. There are great differences on how different countries and/or authorities conduct crash analysis however, there is a common structure in all reviewed methodologies, consisting of four steps.

Data availability as well as the quality of the data are strong determinants of the level of detail of the analysis. Crash data is not always correctly/accurately geolocated while there is some uncertainty to injury severity classification. The use of more advance approaches such as crash prediction models or multiple classification criteria are less commonly met, due to their complexity.

Future research should explore the feasibility of having a unified approach for the assessment of roads across all European countries. It is also important to start shifting the focus towards proactive safety assessment methodologies, that do not rely in crash data and so, is not affected by the limitations related to crash data.

References

1. American Association of Highway Transportation Officials (AASHTO). Highway safety manual, 1st Edition. 2010.
2. Szénási, S. and Jankó, D., 2017. A method to identify black spot candidates in built-up areas. Journal of Transportation Safety & Security, 9(1), pp.20-44.
3. Lam, W.W., Yao, S. and Loo, B.P., 2014. Pedestrian exposure measures: a time-space framework. Travel behaviour and society, 1(1), pp.22-30.
4. Papadimitriou, E., 2016. Towards an integrated approach of pedestrian behaviour and exposure. Crash Analysis & Prevention, 92, pp.139-152.
5. Fournier, N., Christofa, E. and Knodler Jr, M.A., 2019. A mixed methods investigation of bicycle exposure in crash rates. Crash Analysis & Prevention, 130, pp.54-61.
6. Wijnen, W., Weijermars, W., Schoeters, A., van den Berghe, W., Bauer, R., Carnis, L., Elvik, R. and Martensen, H., 2019. An analysis of official road crash cost estimates in European countries. Safety science, 113, pp.318-327.
7. ASFINAG. 2011. Methodology for road safety assessment, "Network Safety Management (NSM) – primary road network Austria.
8. Ghadi, M. and Török, Á., 2017. Comparison different black spot identification methods. Transportation research procedia, 27, pp.1105-1112.
9. Brannolte U., Munch A., Vo H. 2009. "Software-based road safety analysis in Germany", available at: http://www.internationaltransportforum.org/irtad/pdf/seoul/6-Brannolte.pdf
10. African Development Bank Group. 2014. "Existing Roads: Reactive Approaches", available at: https://www.afdb.org/fileadmin/uploads/afdb/Documents/Publications/ROAD_SAFETY_MANUALS_FOR_AFRICA_%E2%80%93_Existing_Roads_Reactive_Approaches.pdf
11. LIETUVOS RESPUBLIKOS SUSISIEKIMO MINISTRO. 2011. "Identification of critical sections on main roads", available at: https://eseimas.lrs.lt/portal/legalAct/lt/TAD/TAIS.401193?jfwid=mmceomu7t
12. Nguyen, H.H., Taneerananon, P. and Luathep, P., 2016. Approach to identifying black spots based on potential saving in crash costs. Engineering journal, 20(2), pp.109-122.
13. Cheng, W. and Washington, S.P., 2005. Experimental evaluation of hotspot identification methods. Crash Analysis & Prevention, 37(5), pp.870-881.
14. Turner, S., Roozenburg, A., and Francis, T. 2006 Predicting Crash Rates for Cyclists and Pedestrians. In Land Transport New Zealand Research Report 289.