TRANSFERRED VERSUS LOCAL SAFETY PERFORMANCE FUNCTIONS: A GEOGRAPHICAL ANALYSIS CONSIDERING TWO EUROPEAN CASE STUDIES.

by Paolo Intini[1], Nicola Berloco[1], Rita Binetti[1,2], Achille Fonzone[2], Vittorio Ranieri[1], Pasquale Colonna[1]

[1]Department of Civil, Environmental, Building Engineering, and Chemistry, Polytechnic University of Bari, 70125 Bari, Italy.
[2]Transport Research Institute, Edinburgh Napier University, EH141DJ Edinburgh, United Kingdom.

ABSTRACT

Two main approaches can be used to predict road accidents: transferring existing Safety Performance Functions (SPFs) from other areas (transferred SPFs), and developing local SPFs. Both approaches have advantages and disadvantages, and are affected by the difficult choice of predictors. Regional variables or terrain factors may lead prediction improvements. However, results from previous relevant research are contradictory and transferability assessments are mainly based on North-American experiences.

Because of these inconsistencies, this study is an attempt of providing new insights on the choice between alternative accident prediction methods by taking into account the geographic variability in the European context. In particular, it addresses three main issues: 1) it compares the prediction accuracy of transferred and local SPFs; 2) it determines the significance of regional factors in explaining safety performances, 3) it assesses the variability of results among the different contexts considered. Research questions are addressed as based on two-lane rural road sites in Italy and Scotland.

The analysis shows differences between the two countries, due to the different nature of the networks, but not within each country. Both advantages and disadvantages were highlighted in the evaluation of transferred and local SPFs. Calibration of transferred SPFs may be less demanding than their local estimation, even if they may lead to unreliable estimates when compared to comprehensive SPFs. However, locally developed SPFs may not provide more significantly reliable estimates than transferred SPFs. Segment curvature and shoulder types are statistically significant predictors in both the Italian and Scottish models, even having different importance.

KEYWORDS: Safety Performance Functions, Transferability, Highway Safety Manual, Regional variables, Two-lane Rural Roads

1. INTRODUCTION

The advances in road safety research can assist practitioners in making technical choices. In particular, the road safety practice may benefit from quantitative predictions of crash occurrence. The use of Safety Performance Functions (SPFs) and Crash Modification Factors (CMFs) greatly helped in making quantitative estimates (see e.g. Hauer and Persaud, 1997; Hauer, 1999; Hauer et al., 2012).

A Safety Performance Function (SPF) is a regression model which links the crash frequency (and/or severity) to predictor variables, usually road and traffic features (AASHTO, 2010). It is developed for different road types, i.e. segments or intersections of rural or urban highways/freeways. Crash Modification Factors (CMFs) (or functions) are factors/functions that account for the effect of a change in some default road conditions (change in road geometric characteristics or traffic control systems) on the accident frequency. They can be applied to the results obtained from a SPF to account for differences with respect to the SPF base conditions. SPFs were taken into account in this article since they consider the influence of different variables on accidents through a single model and thus are used for making predictions.

However, the transferability of SPFs developed in given geographic areas to other countries/areas, may be unfeasible to some extent (see e.g. Sacchi et al., 2012, Farid et al., 2018b). Differences in road contexts, drivers' populations and behaviour, crash database, may result in unreliable transferability of functions to other contexts (see e.g. Bahar and Hauer, 2014; Farid et al., 2016).

1.1 **Background on transferability of accident predictive methods**

Two main strategies may be used to overcome the transferability issue.

The first strategy consists in transferring SPFs from other areas (Transferred Functions, TFs), and calibrating them by correcting their outcomes according to local conditions. A possible basic calibration method is provided in the Highway Safety Manual (HSM) (AASHTO, 2010). Local calibration factors are computed as the ratio of the crashes observed on a sample of local road sites; to those predicted by the base model for the same types of sites. However, a single calibration factor could not be sufficient for large/not homogeneous areas (e.g. different terrains) (Bahar and Hauer, 2014). Hence, different calibration factors may be achieved in case of different local characteristics (see e.g. Tarko, 2006; Colonna et al., 2016a). More refined calibration techniques were defined, which may provide more reliable estimates. For example, the calibration of model parameters through maximum likelihood estimation (Sawalha and Sayed, 2006); segment-specific calibration (Farid et al., 2016); calibration functions (Srinivasan et al., 2016); calibration based on local regression (Farid et al., 2018b) or on the k nearest neighbour data mining method (Farid et al., 2018a), were proposed.

The second strategy consists of developing a local SPF (Local Function, LF) based on data related to the same local road sites. The number and type of independent variables may be the same, or they may be locally adapted, according to the relevant road features in the network. For example, while developing LFs for the Utah State, Brimley et al. (2012) included the multiple-unit trucks traffic percentage as a variable, usually not considered in other studies. Gooch et al. (2018) highlighted that separate predictions can be made for curved segments and tangent sections. Moreover, the choice of the SPF functional form may also be based on the best fitting model. For example, Farid et al. (2019) tested several possible different SPF modelling techniques, by assessing their outcomes and advantages in different conditions. An extended review of possible alternative methods for modelling crash frequency data, together with their assessment, was provided by Lord and Mannering (2010).

However, the choice between these strategies is not straightforward. In fact, while the estimation of LFs is generally encouraged (see e.g. AASHTO, 2010), it could require more resources than simple TF calibration, especially for practitioners. Benefit-cost evaluations could be used to assess if a LF is really needed compared with calibration of a TF, and if its cost may be justified. However, even if there are cases in which the lack of necessary and quality data (see e.g. Gomes et al., 2019) may discourage from trying estimating SPFs; knowing in advance if the LF will outperform results from calibration of TFs is hard, even in presence of reliable and abundant data. On the other hand, there are cases in which the transferability of SPFs can be possible. This may depend on the quality of the reference SPF (Persaud et al., 2002), on the differences between the two areas on which SPFs are developed and transferred (see e.g. Farid et al., 2016), or on modelling techniques (Farid et al., 2019).

1.2 **Background on the geographic variability of the transferability issues**

The transferability issue gets more complex if the variability of the geographic spatial resolution is considered. In fact, defining 1) the boundaries of the areas within which the performed calibration of a transferred SPF (TF) is valid, or 2) the boundaries for using a locally developed SPF (LF) in other parts of the same country/state is arduous.

For example, concerning the first point, calibration factors for TFs may greatly vary for different regions of the same country (Colonna et al., 2016a), or even in sub-networks of the same state (Tarko, 2006). However, country-wide calibrations were conducted as well (see e.g. La Torre et al., 2014).

Similarly, for the second stated point, contradictory results were found. Qin et al. (2002) found no statistically significant differences between four US States on crashes predicted through a model including road and traffic variables. Moreover, Farid et al. (2018b) found that in some cases, US state-specific SPFs may be transferred to other US states. Whereas, calibrations were conducted (e.g. Sun et al., 2006; Garber et al., 2010; Xie et al., 2011; Shin et al., 2015) for transferring American HSM SPFs (AASHTO, 2010) to single US States, resulting in some cases in relevant model corrections. Five different SPFs were developed even in a small State (Virginia, USA), accounting for different commuting patterns, driver behaviour, routes, crash statistics, topography (Garber et al., 2010). This approach was also used in Pennsylvania (USA) (Donnell at al. 2014), where a State-wide SPF was locally adjusted, showing significant prediction improvements, especially at the district level. The application of geographically weighted regressions within a single US state (Virginia) successfully led to different LFs accounting for spatial variability of crash predictions as well (Liu et al., 2017).

The same transferability issues found for the US States may be replicated, to some extent, for other countries, even smaller. For example, two SPFs for the Southern Italian two-lane rural road network (Cafiso et al., 2010; Russo et al., 2016) exist. However, an application of these SPFs in the same area (Colonna et al., 2018) revealed that their outcomes may be largely different depending on the application (i.e. assessment of safety measures or predictions in the road design stage). It is important to note that a consistent part of research about SPFs (both estimation and transferability) was conducted in the USA, with some notable exceptions, such as some European studies (see e.g. Yannis et al., 2016). Moreover, apart from jurisdictional variability, other geographic factors may be influential as well, such as terrain. Zegeer et al. (1987) found that single-vehicle accident rates are higher for mountainous/rolling terrains than for flat ones. A different influence of flat, rolling, mountainous terrains on crash occurrence and slight discrepancies between flat and mountainous terrains were revealed by Srinivasan and Carter (2011) and Bauer and Harwood (2000), respectively.

Hence, it is evident how geographic factors (not only jurisdiction-related) may both affect the transferability of SPFs and the development of calibration factors. Recent studies have then focused on considering geographic factors for crash analyses at different levels: i.e. at the provincial level while taking into account macro-variables (Gonzalez et al., 2018), or even more disaggregate levels while considering a mix of macro and local variables (Lee et al., 2017). However, several variables related to road geometry, traffic operations, and boundary conditions should be considered in the SPF estimation (see e.g. Hauer, 2015). Given their consistent importance revealed in previous research (e.g. Abdel-Aty and Radwan, 2000; Greibe, 2003; Cafiso et al., 2010), the assessment of geographic variability should not be conducted independently from other road geometric and traffic variables.

1.3 Research questions

For the reasons explained above, different geographic factors (at least jurisdiction and terrain variability) should be considered while both calibrating TFs and estimating LFs. However, the choice between calibrating TFs and estimating LFs at the local level is not strongly documented in different contexts. In this regard, contradictory results were found in previous literature, and they mostly belong to North America. Thus, this study would provide additional insights in this field, by analysing datasets from two European case studies.

Hence, this article attempts to address the following research questions. They regard both the choice between using different strategies for local crash predictions and the need for considering geographic factors in the European context:

- Are there significant differences between the outcomes of TFs and estimated LFs?
- Among all the other variables, are geographic factors significant variables for crash predictions, by using both TF calibration and LF development techniques?
- Are the answers to the questions above variable as well, if different geographic areas are considered?

The above reported questions are specifically addressed through the analysis of two separate European traffic and accident database from Italy and Scotland (United Kingdom). The methods employed for data analysis are presented in next section. Results are then reported and discussed.

Complementary to the research aims, this article provides novel SPFs for Italy and Scotland and calibration coefficients for Scotland, which may be of practical use for analysts and engineers. While previous studies report SPFs for Italian two-lane rural roads (Cafiso et al., 2010; Russo et al., 2016), no similar studies were found for Scotland, to the current authors' knowledge. Hence, the present study is deemed useful for enlarging the global dataset of SPFs too (see e.g. PRACT project).

2. METHODS

The general procedure adopted, the database used, the specific variables considered, the calibration procedure and regression techniques employed are described in detail as follows.

2.1 Procedure

The general procedure adopted in this study is divided into the following subsequent stages:

- Transfer the HSM SPF for two-lane rural roads to both the Italian and Scottish contexts, with different refinements: by determining both a state-wide and more detailed calibration factors;
- Develop LFs for the same sample of Italian and Scottish sites used for HSM calibration;
- Compare the results obtained from TF (HSM SPF) calibration with those from LFs estimation;

- Assess the general influence of geographic variability factors on crash predictions; i.e. if the geographic factors (both different geographic areas and terrains) may influence the calibration factors or be included in the regression analysis;
- Compare the results obtained through the studies performed in Italy and Scotland, by focusing in particular on the comparison between the statistically significant variables of the two LFs, and between the factors which may influence the calibration coefficients of TFs.

A concept map of the above described procedure, including links to the structure of this article, to indicate the sections in which each part of the work is discussed, is provided in Figure 1.

Different SPFs may have been considered for TF calibration for both Italy and Scotland. However, the sequential application of HSM SPF and CMFs for two-lane rural roads includes a wider list of road and traffic accident predictors than several other alternative models. For example, Colonna et al. (2018) highlighted that the two-lane rural HSM SPF calibrated for Italy can account for several road and traffic features, when compared with alternative Italian models (Cafiso et al., 2010; Russo et al., 2016). Thus, the base HSM SPF (and related CMFs) were selected as they may represent a wide range of road and traffic characteristics. Moreover, the HSM SPF represents a usual benchmark TF in previous research (see e.g. Sacchi et al., 2012).

The specific choice for two different European areas such as Italy and Scotland was justified by the following remarks. The European continent has a total area comparable to the United States. Hence, as transferability issues were highlighted within the US country, it is possible that different outcomes could result from different European areas, which in addition are different countries. Hence, two different European contexts were chosen (Italy in the Southern Europe and Scotland in the North-Western Europe), characterized by different extension, population distribution, road infrastructure system development (see Table 2), and rule of the road. The Scottish case study was not extended to the whole United Kingdom, to preserve these differences.

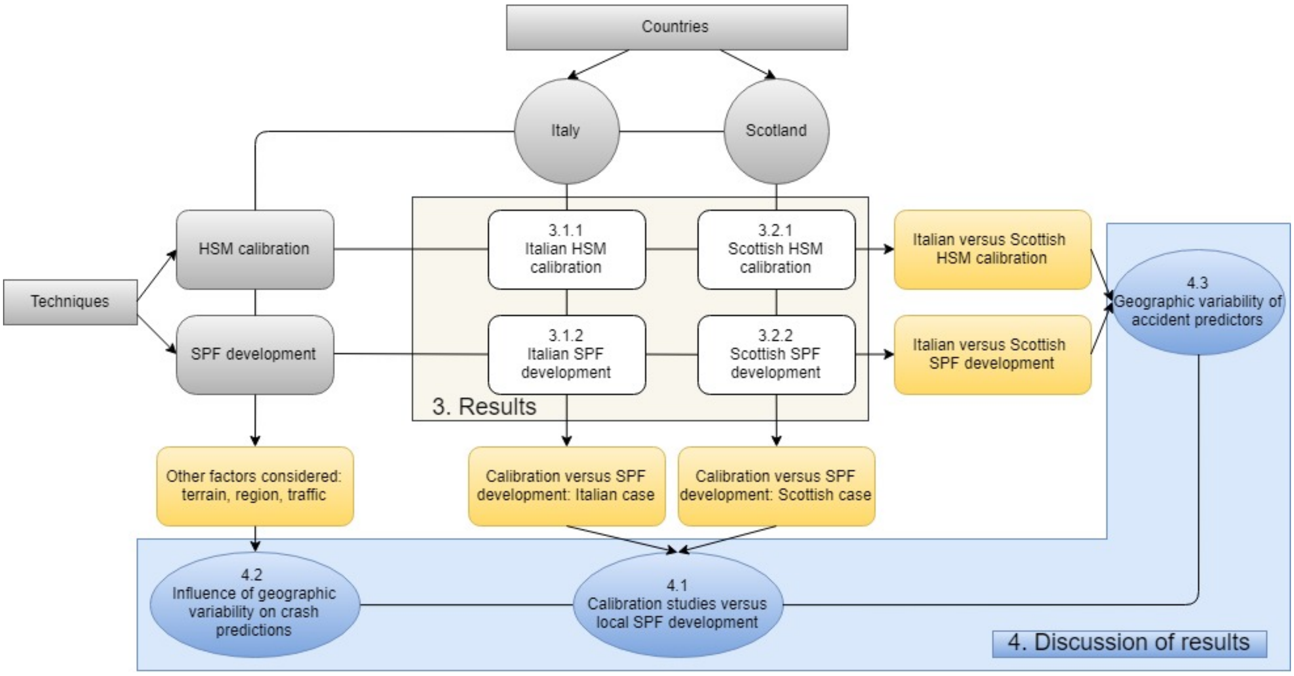

Figure 1. Concept map of the general procedure used, and of the results and discussion sections.

2.2 Database

Two separate databases, namely, for Italy and Scotland, were used. Both database are composed of a first traffic volume dataset, and a second accident (fatal+injury only) dataset for two-lane rural roads. Hence, only the secondary road networks of the two areas were considered, thus excluding roads belonging to the primary and main road networks ("A" and "B" class in the Italian classification, Italian Ministry of Infrastructures and Transport, 2001; motorways and "A" class in the UK classification, UK Department for Transport, 2012). Italian primary and main roads should be designed as multi-lane roads (whether being motorways or not). Whereas, the main UK roads ("A" class) may include also some two-lane roads. However, "A" class roads were not considered in the road network composed of secondary roads, to be coherent with the Italian case.

Annual average daily traffic counts were collected from the respective road agencies (UK Department for Transport, covering all the Scottish network; Italian ANAS, covering part of the Italian network). Accident data were retrieved from different sources: Italian National Institute of Statistics (ISTAT) and Italian Automobile Club (ACI) for the Italian case and the online portal https://data.gov.uk/ for the Scottish case. At least three years of accident data were collected (see Bahar and Hauer, 2014).

Starting from the overall database, traffic and accident data were coupled for road sections provided with traffic counts. A road section is defined here as a section on a road trunk included between two relevant intersections (i.e. with roads of similar importance, excluding driveways or intersections with minor roads), on which a unique traffic volume is assigned, since it is deemed as constant along it. The resulting total length of segments inquired is about 213 km (74 segments) for Italy and 180 km for Scotland (66 segments).

The total number of observed Scottish crashes is low (101 in total), even if the total length of segments investigated is comparable with the Italian one. Hence, among all the segments provided with traffic data, a subset was selected in compliance with both the following requirements: 1) having an equivalent number of at least 100 accidents/year (AASHTO, 2010), 2) including a sufficient number of zero-count sites to account for the low mean estimated accidents/km rate in the part of network investigated. Detailed information concerning the road segments composing the final database obtained are reported in the following Table. Information about the dataset are also classified according to the traffic ranges and regions of the segments, which pertains to the main research questions. Descriptive statistics are also reported about accidents, traffic, geometric and other characteristics of the segments in the dataset. The variables considered in this study are described in detail in 2.3.

Table 1. Descriptive statistics of the variables considered among the sample of segments, showing the mean values (st. dev. in brackets) or counts associated to each variable over the considered road segments (in all the database, for the specific region, for the specific traffic range).

Variables	**Descriptive statistics**				
	Overall	**Region 1**	**Region 2**	**Traffic Range 1**	**Traffic Range 2**
Territory: Italy (years of data: 2007-2012)					
	-	**Northern Italy**	**Central-Southern Italy**	**≤10,000**	**>10,000**
Number of Segments (-)	74	20	54	56	18

Homogeneous sub-segments (Sites) (-)	398	112	286	316	82
Total Length of Segments (km)	212.57	53.82	158.74	163.53	49.03
Total Accidents (accidents)	530	260	270	242	288
Accident Frequency (accidents/year)	1.19 (1.74)	2.17 (2.27)	0.83 (1.35)	0.72 (1.11)	2.67 (2.43)
Accident Frequency per km (accidents/year/km)	0.44 (0.51)	0.84 (0.60)	0.29 (0.37)	0.25 (0.31)	1.03 (0.54)
AADT (vehicles/day)	6506.53 (4269.27)	9927.00 (4811.17)	5239.69 (3279.70)	4484.14 (2410.54)	12798.39 (2019.65)
Length of Segments (m)	287.25 (1700.58)	2690.95 (1661.53)	2939.70 (1725.28)	2920.25 (1678.85)	2723.83 (1807.99)
Road Width (m)	8.83 (1.12)	8.79 (1.11)	8.85 (1.13)	8.77 (1.13)	9.01 (1.07)
Shoulder Type (-) (categorical)	Paved – 30 Gravel - 3 Composite/ Mixed – 25 Turf - 16	Paved – 6 Gravel - 0 Composite/ Mixed – 8 Turf - 6	Paved – 24 Gravel - 3 Composite/ Mixed – 17 Turf - 10	Paved – 22 Gravel - 3 Composite/ Mixed – 19 Turf - 12	Paved – 8 Gravel - 0 Composite/ Mixed – 6 Turf - 4
Radius of Curvature (m)	294.62 (194.73)	275.32 (171.66)	301.86 (204.28)	300.27 (207.91)	269.22 (123.75)
Curve Ratio (-)	0.14 (0.15)	0.14 (0.12)	0.14 (0.16)	0.16 (0.16)	0.08 (0.09)
Slope (%)	2.83 (2.06)	1.78 (1.64)	3.21 (2.08)	3.31 (2.09)	1.33 (0.98)
Driveway Density (driveways/km)	7.53 (14.23)	8.82 (15.08)	7.05 (14.02)	5.78 (9.21)	12.99 (23.52)
RHR (-) (categorical, integers: 1-7)	4.14 (1.16)	3.77 (1.32)	4.27 (1.07)	4.23 (1.09)	3.85 (1.34)
Elevation (-)	Flat – 37 Rolling - 37	Flat – 13 Rolling - 7	Flat – 24 Rolling - 30	Flat – 25 Rolling - 31	Flat – 12 Rolling - 6
Territory: Scotland (years of data: 2012-2014)					
	-	**South (Western/ Eastern) Scotland**	**Highlands-Island/ Eastern Scotland**	**≤2,000**	**>2,000**
Number of Segments (-)	66	43	23	41	25
Homogeneous sub-segments (Sites) (-)	311	203	108	196	115
Total Length of Segments (km)	180.22	117.79	62.43	112.20	68.02
Total Accidents (accidents)	101	59	42	55	46
Accident Frequency (accidents/year)	0.51 (0.63)	0.46 (0.51)	0.61 (0.80)	0.45 (0.44)	0.61 (0.85)
Accident Frequency per km (accidents/year/km)	0.20 (0.32)	0.17 (0.23)	0.27 (0.43)	0.17 (0.20)	0.25 (0.45)
AADT (vehicles/day)	2048.06 (1620.94)	1934.07 (1586.63)	2261.17 (1698.27)	992.07 (444.50)	3779.88 (1325.74)
Length of Segments (m)	2730.62 (1525.36)	2739.30 (1434.47)	2714.39 (1716.27)	2736.51 (1529.61)	2720.96 (1549.78)
Road Width (m)	8.16 (1.53)	8.19 (1.42)	8.11 (1.75)	7.84 (1.39)	8.70 (1.62)
Shoulder Type (-)	Paved - 1	Paved - 1	Paved - 0	Paved - 0	Paved - 1

(categorical)	Composite/ Mixed – 24 Turf - 41	Composite/ Mixed – 14 Turf - 28	Composite/ Mixed – 10 Turf - 13	Composite/ Mixed – 10 Turf - 31	Composite/ Mixed – 14 Turf - 10
Radius of Curvature (m)	348.58 (274.74)	356.61 (318.32)	333.55 (170.90)	276.39 (173.58)	466.96 (361.54)
Curve Ratio (-)	0.55 (0.26)	0.55 (0.25)	0.56 (0.28)	0.51 (0.24)	0.62 (0.27)
Slope (%)	3.34 (1.52)	3.51 (1.48)	3.03 (1.58)	3.57 (1.50)	2.97 (1.51)
Driveway Density (driveways/km)	3.86 (2.35)	3.61 (2.56)	4.35 (1.84)	3.90 (2.66)	3.80 (1.78)
RHR (-) (categorical, integers: 1-7)	5.62 (0.76)	5.76 (0.58)	5.36 (0.97)	5.78 (0.73)	5.40 (0.77)
Elevation (m)	105.47 (63.91)	105.84 (59.67)	104.76 (72.60)	109.31 (67.73)	99.16 (57.88)

2.3 Variables

The independent variables considered for calibrating TFs and developing LFs are here defined and described. Given the research questions, a separate section is dedicated to geographic variables.

2.3.1 Geographic variables

Coherently with the study aims, geographic factors were considered within each country and not only as the difference between countries (i.e. Italy versus Scotland). Hence, both Italy and Scotland were divided into regions, used as synthetic variables to capture the influence of socio-economic and driving behavioural differences. Italy (I) and Scotland (S) are hardly comparable in terms of area (approx. 300,000 km^2 (I) and 80,000 km^2 (S)), population (approx. 60 million inhabitants (I), 5 million inh. (S)). However, both Italy and Scotland were divided into two main regions (see Fig. 2). This was made to avoid excessive fragmentation of the database into several small regional sub-sets not ensuring statistical representation of the area, given also the length of the sample of segments inquired. The considered regions are defined as follows:

- Italy: 1) Northern Italy, 2) Central-Southern Italy;
- Scotland: 1) “Lowlands” (Southern part), 2) “Highlands” (Northern part).

The two Italian macro-regions were chosen based on the EU NUTS 1 level classification (European Parliament and Council, 2003). This classification was deemed useful to reveal regional differences, since it is based on socio-economic features (European Union, Eurostat, 2015). Central Italy (which occupies a limited territory) and Southern Italy were further grouped together, to avoid excessive fragmentation. The obtained two regions (Northern and Central-Southern Italy) have similar populations, but they differ in densities and some other socio-economic variables (see Table 2).

The two Scottish macro-regions were chosen based on the division into Lowlands and Highlands, with historical and socio-cultural roots (e.g. Devine, 1979, Davidson, 2000). The macroscopic EU NUTS 2 level classification (European Parliament and Council, 2003) divides Scotland into 4 regions: East, South-West, North-East, Highlands/Islands. However, Scotland (far less wide than Italy) was divided into two regions as well as Italy. Hence, Eastern and South-Western NUTS regions were grouped into a “Lowlands” macro-region. Since North-Eastern Scotland is small and less densely populated than the other Southern areas, it was grouped with the adjoining Highlands and Islands NUTS region into a “Highlands” macro-region. As can be noted from Table 2, the division of Scotland into Highlands (North) and Lowlands (South), based on traditional historic classifications, is justified by geographic (i.e. population and population density) and infrastructural differences (variable "density of motorways" in Table 2), rather than other socio-economic comparisons.

Table 2. Geographic and socio-economic variables for Italy and Scotland (data source: http://ec.europa.eu/eurostat/data/database).

Variables	**Italy**		**Scotland**	
	Northern Italy	**Southern and Central Italy**	**"Highlands" (Highlands/ Islands and North-Eastern Scotland)**	**"Lowlands" (Eastern and South-Western Scotland)**
Population (millions)[1]	30.94	22.40	0.97	4.43
Area (km^2)	120,260	131,275	48,518	31,715
Density (inhabitants/km^2)[1]	257.32	170.61	19.90	139.82
Gross Domestic Product per 1000 inhabitants[2] [€]	32.63	22.34	42.47	33.45
Rate of long-term unemployment (≥ 12 months) with respect to active population[3] [%]	3.78	8.92	2.93[4]	2.73
Life expectancy[2] [years]	83.39	82.94	80.33	79.80
Intentional homicides per 100 inhabitants[5]	0.06	0.15	-	-
Density of motorways[6] [m/km^2]	33.92	16.76	0.00	18.95[7]

[1]as of 2017; [2]average on the period: 2014-2016; [3]average on the period: 2012-2014; [4]Including only Highlands and Islands region; [5]average on the period: 2008-2010; [6]as of 2015; [7]based on Transport Scotland (2016).

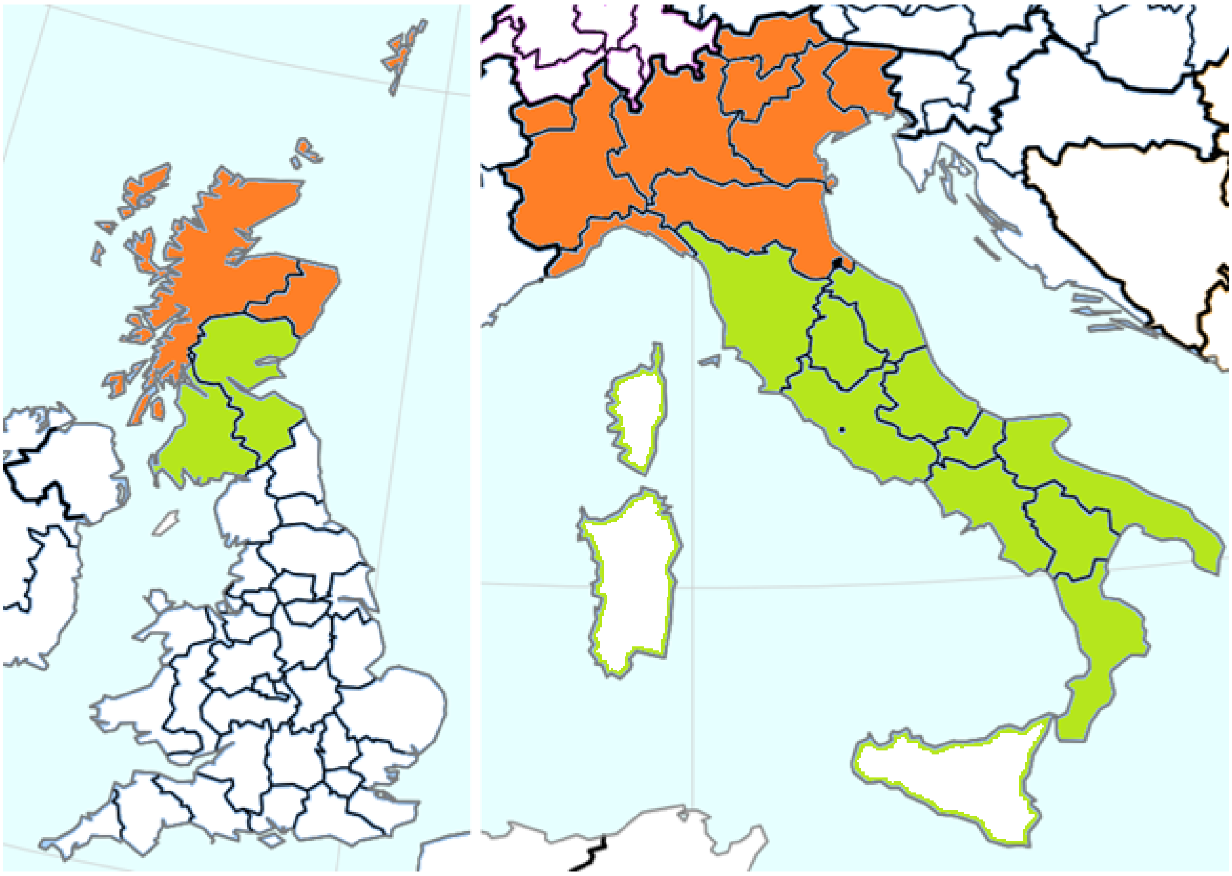

Figure 2. Map of regions considered in this study. Left: Scotland ("Highlands" in orange; "Lowland" in green). Right: Italy (Northern Italy in orange; Central/Southern Italy in green). Based on: http://ec.europa.eu/eurostat/web/nuts/nuts-maps-.pdf-.

Both the above highlighted intrinsic differences between countries (Italy/Scotland) and within countries (different regions) are helpful for the aims of this study. In fact, they are useful to assess if both the methods for safety predictions: calibration of TFs and estimation of LFs, may be universally applied or they are dependent on: 1) the specific area considered, 2) its inner regional variability.

Apart from regional boundaries, also terrain type was considered in this study, as it may influence accident prediction (Carter and Srinivasan, 2011; Bahar and Hauer, 2014).

For the Italian dataset, road sites were classified into: flat and rolling terrain (the latter is the most widespread in Italy) (Colonna et al., 2016a). In the cited study, a binary terrain class (flat or rolling) was assigned to each road site according to the average terrain elevation above/below the site. Mountainous terrains were not present in the database. The elevation threshold between flat and rolling terrains was set to 400 m above mean sea level. This value was previously identified as an indicative limit beyond which the alignments of the secondary roads inquired are highly influenced by surrounding terrains, through exploration of the road segments in the sample (Colonna et al., 2016a). In this regard, the differences between the average gradients of segments and their variation within the segment are shown in Figure 3. Boxplots clearly show how the two populations of gradients above and below the 400 m selected threshold are different. Vertical alignments are more varying and gradients are significantly steeper in the "rolling" than in the "flat" terrain class.

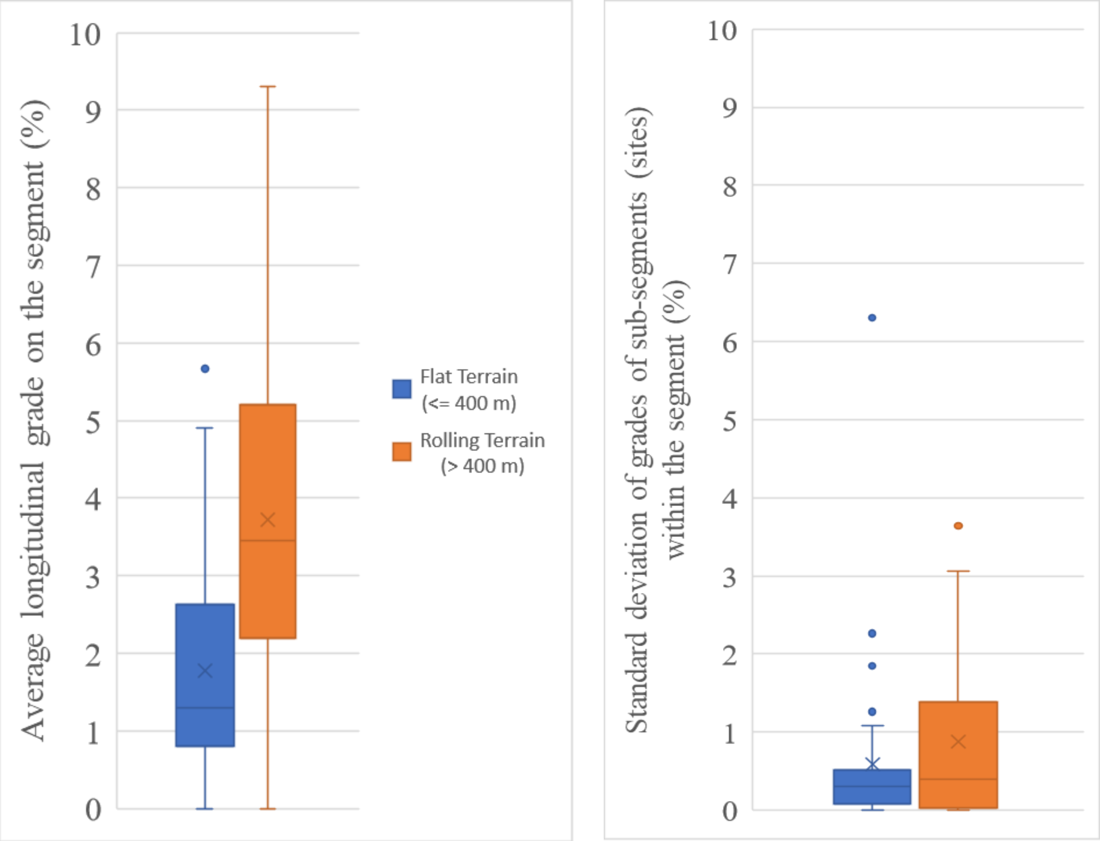

Figure 3. Boxplots of: (left) the average longitudinal grades on the Italian segments, (right) standard deviation of grades of Italian sites (sub-segments) within segments; on "flat" and "rolling" terrains.

For the Scottish dataset, the average terrain elevation (m) collected for each road site, revealed an overall distribution of elevations far below 400 m. Hence, in the Scottish case, no variability due to terrain was inquired.

2.3.2 Other variables

Apart from geographic variables (region, terrain), the other variables used in this study are the several predictors included in the HSM (AASHTO, 2010), both in the base SPF and related CMFs.

Except from traffic data provided by road agencies, road-related information were manually retrieved by using different software applications, since reliable geometric inventories were scarce or absent. Most information were collected through Google Earth© and Google Street View©, coherently with some other previous applications (e.g. La Torre et al., 2014; Shin et al., 2015).

The variables: AADT, length of sites, road width, shoulder type, radius of curvature, presence of Two-Way Left Turn-Lanes (TWLTL), are deemed as necessary for calibrating a TF, while other variables are indicated as only desirable (AASHTO, 2010). However, since the aims of this study include also LF estimation, then information concerning also desirable variables were collected. No segments with automated speed enforcement, centerline rumble strips were found in the two database, and no segments with road lighting, passing lanes and TWLTL (right turn-lanes in the Scottish case) were found in the Scottish database (only few in the Italian one). For this reason, those variables were not further considered for SPF development. Moreover, the variable: variance of superelevation at horizontal curves (with respect to the one prescribed) was excluded due to unreliable measures achievable through the applications used for data collection mentioned above. The rating variable: Roadside Hazard Rating -RHR- was assigned by visually checking the on-site conditions and comparing them to the illustrative conditions indicated in the HSM (AASHTO, 2010). Details concerning the variables taken into account: AADT, length, road width, shoulder type, radius of curvature, slope, driveway density, are reported in Table 1.

The road sections (between two major intersections or significant cross-sectional changes) included in the database may have a significant length (between 2.5 and 3 km on average, see Table 1). Hence, they are generally composed of sub-sections (sites) having different characteristics (e.g. presence of curves, changes in slopes, shoulder widths, etc.). Each site composing the whole section is defined as being internally geometrically homogeneous (i.e. all the variables taken into account do not significantly change among it). Due to the noticeable length of most sections in the database, the total length of sites collected on different parts of the section (henceforth referred to as segment length) is not equal to the total section length. The "segment" is then composed of different homogeneous sites (e.g. hs-1, hs-2, hs-5, etc., see Fig. 4).

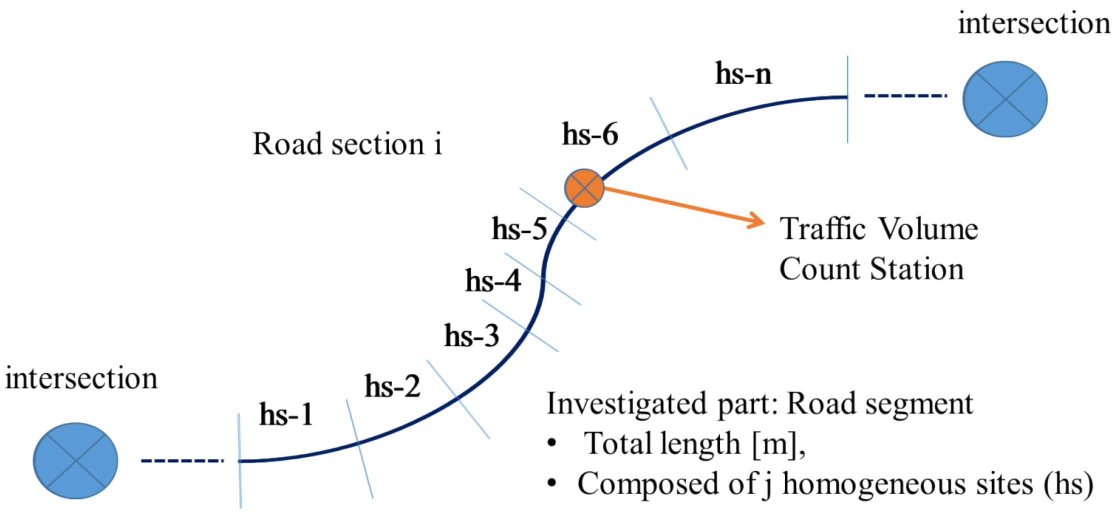

Figure 4. Graphical scheme of road section and homogeneous sites.

The variables: road width, radius of curvature, slope and RHR may then have different values for each site along the road segment. Hence, for each of the variables listed above, an average value weighted according to the road site lengths, is then computed and assigned to each road segment.

To provide indications concerning the average curvature of each road segment investigated, the variable “Curve Ratio” (Cafiso et al., 2010) was computed, by dividing the total length of curved sub-segments by the total segment length. The variable “Shoulder Type” may univocally be assigned to each homogeneous site, if right and left shoulders are similar. In case of right shoulders different from the left ones, or shoulder type varying along the road segment, “Shoulder Type” is set to ‘mixed’, and aggregated to the modality ‘composite’, since different materials are combined.

2.4 Calibration procedure

The performed calibration of a transferred SPF (TF) adopts: 1) the HSM (AASHTO, 2010) model for two-lane rural roads as base reference SPF; 2) the calibration procedure described in the HSM for transferring SPFs to different jurisdictions, (considering also improvements proposed by Lord et al., 2016); 3) a procedure aimed at assessing the reliability of calibration (Bahar and Hauer, 2014).

The unit of reference for calibration is the homogeneous road sub-segment (site), to which a set of parameters should be univocally assigned. The HSM indicates that a reliable calibration should be at least based on:

- 30-50 homogeneous road sites;
- 100 accidents/year over the total sample of sites;
- 3 recent years of accident data.

The minimum number of segments is respected for each subset considered (two regions and traffic ranges for each territory). The requirement concerning the minimum years of data was met for both the Italian (6 years) and Scottish (3 years) database. The Italian calibration was limited to 5 years of data (coherently with other studies, e.g. La Torre et al., 2014), since long periods are discouraged for calibration studies. In fact, calibration factors may vary over time.

For what concerns the minimum number of accidents, these are total accidents. Since fatal and injury accident data are often more reliable than total accident data (or the only available), a sample composed of a number slightly minor than 100 fatal+injury accidents per year may be sufficient (e.g. Sacchi et al., 2012). The Italian database is composed of 422 fatal injury accidents over the period 2008-2012 (84.4 fatal+injury accidents/year). Hence, the requirement is deemed to be met for the Italian case, and not for the Scottish case (101 fatal+injury accidents in the period: 2012-2014, 33.7 fatal+injury accidents/year). However, based on the information included in the accident database investigated and their descriptions, the fatal+injury Italian and Scottish were equated to, namely, KAB accidents (Colonna et al., 2016a; Cafiso et al., 2012) and KABC accidents (which account namely for about 18 % and 32 % of total accidents, according to HSM estimates). The reference scale taken into account is the KABCO scale (K = Killed, A = Incapacitating injury, B = Non incapacitating injury, C = Possible Injury, O = Property Damage Only, PDO), provided in the HSM (AASHTO, 2010). This means that the Scottish 101 fatal+injury accidents may correspond to 316 total accidents, which could meet the HSM recommendations. However, given this uncertainty, which broadly affects the significance of results obtained for specific subsets (regions and traffic ranges), the reliability assessment of calibration results is fundamental.

The calibration procedure was firstly run for the entire dataset, i.e. for estimating single Italian and Scottish calibration factors. Thereafter, the same procedure was repeated by considering different subsets of data for obtaining more detailed calibration factors (Bahar and Hauer, 2014). Given the aims of this article, the above defined regions were used to classify data into regional clusters for calibration purposes. The influence of the traffic volume variability was considered as well to define subsets of data. This choice is based on the nature of the HSM SPF used as reference. In fact, according to this function, the accident frequency on two-lane rural roads is linearly dependent on traffic volume. Since traffic volume is a strongly influential variable on accident frequency (AASHTO, 2010; Greibe, 2003, Abdel-Aty and Radwan, 2000), and the traffic-accidents relationship may also be non-linear (e.g. Kononov et al., 2003), the variability of calibration factors for different traffic ranges was investigated. If calibration factors for different traffic ranges largely differ, then a non-linear traffic-accidents relationship may have been revealed.

For the Italian dataset, 10,000 vehicles/day was identified as a threshold dividing traffic ranges (Colonna et al., 2016a). In fact, previous studies (Sacchi et al., 2012; La Torre et al., 2014) highlighted that the HSM SPF tends to underestimate crash frequencies for high-crash sites, roughly for AADT > 10,000. Whereas, the Scottish dataset is mainly composed of low-volume roads (mean AADT: approx. 2,000 vehi./day, and standard deviation comparable to the mean). Hence, due to the high differences in traffic volumes of the two samples, the same Italian threshold was not deemed usable. Hence, it was set to 2,000 vehi./day; as this is close to the mean value of the sample of segments. In this way, the variability of calibration factors with traffic was investigated for Scotland as well, in the range of the traffic volumes in the sample.

The calibration output is a calibration factor C_x, obtained by dividing the total observed accidents (in this case fatal+injury accidents) on the considered segments by the predicted accidents on the same segments (through the application of the base HSM SPF, the appropriate percentage of accident severities, and the applicable CMFs to each segment, according to the collected variables):

$$C_x = \frac{\sum_{all\ segments} Observed\ number\ of\ accidents}{\sum_{all\ segments} Predicted\ number\ of\ accidents\ (uncalibrated\ SPF)} \quad (1)$$

The calibration procedure was applied for both Italy and Scotland, and for the different subsets considered (two regions and traffic ranges for each country). Hence, an overall factor and other specific calibration factors are obtained.

The C_x factors obtained were assessed by using the approach proposed by Bahar and Hauer (2014). The reliability assessment is based on $cv\{C_x\}$ values (Bahar and Hauer, 2014), computed as follows. They represent an estimate of the coefficient of variation of the associated Cx factors:

$$cv\{C_x\} = \frac{\sqrt{\sum_{j=1}^{n} N_{a,j} + \ k_j N_{a,j}^2}}{C_x \ (\sum_{j=1}^{n} N_{u,j})} \tag{2}$$

Where:
$N_{u,j}$ = uncalibrated predicted number of crashes for the segment j;
$N_{a,j} = \ C_x \ N_{u,j}$ = calibrated predicted crashes for the segment j (replaceable by observed crashes);
k_j = over-dispersion parameter (indicating a variance greater than the mean) of the base HSM SPF.

Values of $cv\{C_x\}$ less than 0.20 may be related to accurate Cx estimates (Bahar and Hauer, 2014). Hence, this value is deemed as a good threshold for assessing the reliability of calibration factors.

The improved guidelines for HSM calibration studies (Lord et al., 2016) were also taken into account, which provide the minimum number of road sites for obtaining a given level of accuracy. This number depends on the coefficient of variation of the observed accidents in the sample. If this minimum number is not achieved at a sufficient confidence level, the LF estimation is advised. Moreover, the need for region-specific calibration factors is suggested as well when the following disequation is satisfied. Otherwise, the State-wide calibration factor may be deemed as sufficient.

,

$$e_r = \left| \frac{\frac{N_{obs,R}}{L_{average,R} * N_{SPF,HSM}(AADT_{average,R})}}{\frac{N_{obs,S}}{L_{average,S} * N_{SPF,HSM}(AADT_{average,S})}} - 1 \right| > 0.10 \tag{3}$$

Where:
$N_{obs,R/S}$ = observed accidents in the Regional (R)/State-wide (S) sample of road sites;
$N_{SPF,HSM}$ ($AADT_{average,R/S}$) = accidents predicted from the baseline HSM SPF as a function of the AADT over the Regional (R)/State-wide (S) sample of road sites;
$L_{average,R/S}$ = average segment length in the Regional (R)/State-wide (S) sample of road sites (km).

Alternative recent approaches may have been used for the HSM calibration (see e.g. Srinivasan et al., 2016; Farid et al., 2018a,b). However, a simple calibration approach was preferred (AASHTO, 2010), to better stress the different predictive capabilities, if any, of two extreme alternatives: LF estimation and TF calibration. However, guidance from Bahar and Hauer (2014) and Lord et al. (2016) were taken into account, as previously indicated, to assess the results from the HSM calibration. Additional references for these selected criteria can be found in Geedipally et al. (2017); Shirazi et al. (2016a,b).

2.5 Modelling techniques

Accident modelling is often conducted by applying General Linear Modelling (GLM) approaches (Lord and Mannering, 2010), more flexible than linear modelling. Accident counts resulted over-dispersed (variance greater than the mean), thus the GLM regression was conducted by assuming a Negative Binomial (NB) distribution of the errors, and a natural logarithmic link function (Hilbe, 2011; Chatterjee and Simonoff, 2013). This approach is commonly used for developing LFs (see Lord and Mannering, 2010 for a list of studies) and specifically for two-way two-lane rural roads (e.g. Zhang and Ivan, 2005; Cafiso et al., 2010; Russo et al., 2016). Zero-inflated models could be also used in these cases, since accident counts are often widely populated of zeros. However, their application was criticized for highway safety purposes (see Lord et al., 2005b) and the percentages of zeros in the sample of yearly accident frequencies are about 50 % (Italy) and 60 % (Scotland).

The open-source software R was used for modelling and statistical analyses, by using the 'MASS' library (Venables and Ripley, 2002). In this package, the over-dispersion parameter of the NB GLM model is estimated through maximum likelihood estimation, which is indicated as the most reliable technique among different possible estimates in the study by Lord (2006).

The chosen model form used for both the Italian and Scottish regressions is expressed as follows:

$$E(Y) = \exp\left(\beta_0\right) * L^{\beta_1} * AADT^{\beta_2} * \exp\left(\sum_{i=3}^{n} \beta_i X_i\right) \tag{4}$$

Where:

E(Y) = predicted number of (fatal+injury) accidents per year (accidents/year);
L = length of the segment (m);
$\beta_0, \beta_2, \ldots, \beta_n$ = estimated coefficients of the regression ($\beta_1$is set to 1);
$X_3, X_4, \ldots, X_n$ = regression variables considered, other than segment length and AADT: road width, shoulder type, radius of curvature, curve ratio, slope, driveway density, RHR, region, elevation.

The *n* variables considered for the regression are the same required for the HSM SPF calibration. The coefficient of the segment length (β_1) was set to 1, as in most of accident prediction models (e.g. Lord et al., 2005a; AASHTO, 2010; Cafiso et al., 2010; Russo et al., 2016), implying a linear relation between segment length and accidents. The variables "right shoulder width", "left shoulder width" and "lane width" were aggregated into a comprehensive variable "road width" (Cafiso et al., 2010), since they are strongly inter-related. In fact, the widths of left and right shoulders are mostly similar, and the widths of lanes and shoulders may both increase with the road importance. The classification of shoulder types into paved, gravel, composite, turf, was further aggregated as well according to the lack and/or scarcity of some shoulder types in the database. In the Italian case, gravel shoulders were aggregated to the composite/mixed ones, due to their scarcity (only 3 segments), thus having only three classes. In the Scottish case, there were no segments with gravel shoulders and only one with paved shoulders. Thus, only two classes were considered: paved/mixed/composite, and turf shoulders, by mixing classes with close effects on safety according to HSM CMFs (AASHTO, 2010).

The variables "Curve Ratio (CR)" and "Radius of curvature" are associated due to their intrinsic definition (the average radius of curves on the segments is finite only if CR ≠ 0). Hence, in order to keep both information by avoiding collinearity, another continuous variable was defined:

$$MC = \left(\frac{1}{MR}\right)_{curved\ part} * CR + \left(\frac{1}{MR}\right)_{straight\ part} * (1 - CR) = \left(\frac{1}{MR}\right)_{curved\ part} * CR \tag{5}$$

Where:
MC = weighted mean of the segment curvature (1/km), equal to zero for straight segments;
MR = mean radius of curvature of the curved part of the road segment (km), set to infinity in straight parts of segments, thus leading to eliminate the second term of the weighted mean.

The list of variables and their nature is summarized in Table 3.

Table 3 – Predictors considered for the SPF development

Variable	Symbol	Type	Unit or Values
Annual Average Daily Traffic volume	AADT	Continuous	vehicles/day
Segment length	L	Continuous	m
Total road width	TW	Continuous	m
Shoulder type	ST	Nominal	*Italy*: 0 – Paved, 1 – Mixed-Composite/Gravel, 2 – Turf *Scotland*: 0 – Mixed-Composite/Paved, 1 – Turf
Weighted mean curvature	MC	Continuous	1/km
Longitudinal slope	i	Continuous	%
Driveway Density	DD	Continuous	Driveways/km
Roadside Hazard Rating	RHR	Ordinal	Range: [1, 7] (only integers)
Region	REG	Nominal	*Italy*: 0 – North, 1 – Centre-South *Scotland*: 0 – Lowlands, 1 – Highlands
Elevation	ELE	Nominal (Italy only)	0 – Flat, 1 – Rolling

Three goodness-of-fit measures related to GLM modelling (see e.g. McCullagh, 1984, or Myers et al., 2012) were used in this study: the AIC (Akaike Information Criterion), the Pearson χ^2 (5 % significance level), and the Nagelkerke pseudo-R^2 (adjusted for non-linear regressions, variable between 0 and 1). The latter two measures can provide information about the goodness-of-fit of each single model developed, while the AIC criterion is useful for comparisons between estimated models. Plots of cumulate residuals (CURE plots) (see Hauer and Bamfo, 1997) were also used to examine the goodness of fit of the estimated models, with specific reference to each included variable.

Among all the possible models obtainable by combining the 10 variables considered, the model showing: 1) the highest goodness-of-fit measures and 2) the highest number of variables for which the estimated parameter is statistically significant at the 90 % confidence level (used in previous similar studies for relatively small datasets, such as Gomes et al., 2012; Oh et al., 2006), was selected.

3. RESULTS

Results of both HSM SPF calibration and SPF development are shown in this section.

3.1 Italian case study

3.1.1 Italian HSM Calibration

Results from the HSM SPF Italian calibration study (updated from Colonna et al., 2016a) are reported as follows, including the assessment measure: $cv\{C_x\}$, and classified according to traffic and regions.

Table 4 – Results of the HSM SPF calibration - Italy

Variable: Region	**AADT Ranges**	**Number of sites**	**Cx**	**cv[Cx]**	**Need for regional Cx (e_r) (Lord et al.. 2016)**
Italy	**Overall**	398*	1.44	0.08	-
	< 10,000	316	1.19	0.10	-
	≥ 10,000	82*	1.75	0.14	-
Northern	**Overall**	112*	1.66	0.15	Yes (0.23)

Italy	**< 10,000**	51	1.39^	0.22	-
	≥ 10,000	61*	1.73	0.17	-
Central-Southern Italy	**Overall**	286	1.29	0.09	Yes (0.13)
	< 10,000	265	1.16	0.11	-
	≥ 10,000	21	1.81^	0.21	-

Note: Cx coefficients marked with the superscript "^" are deemed less reliable due to either related number of segments < 30 or cv[Cx] ≥ 0.20. Numbers of segments marked with the superscript "" are those representing the more reliable subsets for calibration, associated to estimated "confidence levels" (based on Lord et al., 2016) around 70 %.*

All calibration coefficients in Table 4 are reliable (Bahar and Hauer, 2014), except for low traffic in Northern Italy and for high traffic in Southern/Central Italy. For some coefficients, including the overall factor, the estimated equivalent "confidence levels" (Lord et al., 2016) are around 70 %, based on the number of segments in the sample. This may justify HSM calibration instead of SPF development. In particular, regional coefficients are advised for both the macro-regions considered (e_r values > threshold indicated in Eq. 3), especially for Northern Italy (e_r= 0.23).

The HSM SPF generally underestimates accident frequencies for Italian two-lane rural roads (all Cx factors are > 1). There is a notable difference between traffic ranges: Cx considerably higher for high traffic volumes (> 10,000) than low volumes. This result is valid nationwide and even disaggregating data over regions. However, the high difference between Cx values for different traffic ranges for both Northern and Centre-South Italy is not enough reliable due to the associated borderline $cv\{C_x\}$ values. Some reliable Cx factors showing very low $cv\{C_x\}$ values are those obtained for low traffic volumes (nationwide: 1.19, Centre-South Italy: 1.16).

A regional effect can be noted in the outputs of HSM calibration. The overall factor for Northern Italy (Cx = 1.66) is considerably higher than for Centre-South Italy (Cx = 1.29), and indeed a regional calibration factor was deemed necessary based on Lord et al. (2016). However, this difference may be attributed to the high percentage of high traffic sites for Northern Italy, considerably higher than the respective sites for Centre-South Italy. The higher traffic volumes for Northern Italian sites may have led to the notably high Cx for Northern Italy. Hence, pairwise comparisons between regions should be made by differentiating for traffic ranges. When comparing low traffic ranges, a notable difference emerges between Northern (Cx = 1.39) and Centre-South Italy (Cx = 1.16). However, the reliability of the Northern Italian low-volume coefficient is deemed questionable ($cv\{C_x\}$ = 0.22). Whereas, when comparing high traffic ranges, no consistent differences may be noted (North: Cx = 1.73; Centre-South: Cx = 1.81).

3.1.2 Local Safety Performance Function: Italy

The statistical parameters related to the fitted Italian SPF are presented in Table 5, including the over-dispersion parameter □. The NB model satisfactorily fits accident data, by considering the goodness-of-fit measures (in particular the pseudo-R^2).

Table 5 – NB model parameters and goodness of fit measures for the Italian SPF, with p-values and standard errors in brackets

Model	**Parameters**					**Goodness-of-fit**			**Over-dispersion**
	β_0	β_{AADT}	$\beta_{ST=1}$	$\beta_{ST=2}$	β_{MC}	**AIC**	χ^2	$_{Pseudo}R^2$	□

IT	-20.998 (<.001, 0.940)	1.423 (<.001, 0.102)	0.660 (<.001, 0.133)	0.880 (<.001, 0.162)	0.223 (<.001, 0.058)	1078.1	471.9 (0.866)	0.622	3.670 (1.100)

The variables included in the model are: AADT, shoulder type, weighted curvature. They are all significant at the chosen significance level ($p = 0.10$), actually exceeding the 99 % confidence level. As expected, AADT is positively related to the accident frequency, and β is > 1, indicating a more than linear traffic-accident relationship. Coefficients of gravel, composite, mixed (ST = 1) or turf (ST = 2) shoulders are positive, which means that they seem less safe than paved shoulders (reference condition: ST = 0). Weighted mean curvature (MC) is positively related to the accident frequency: the more curved segments on the section and the more the curvature, the higher seems the accident frequency.

Whereas, the following variables did not result statistically significant in the model development at the chosen significance level ($p = 0.10$): total road width, longitudinal slopes, driveway density, RHR, elevation. The variable region resulted statistically significant at the defined confidence level in the alternative model IT(A) reported in Table 6 indeed (as well as the longitudinal slope i, positively related to the accident frequency). However, the AIC value associated to the model IT(A) is greater than the corresponding value in Table 5 and thus, for this reason, the latter model was selected. However, given the research questions of this article, it is important to note that region may be considered as a significant variable in an alternative accident prediction model. Taking into account the model IT(A) in Table 6, Central-Southern Italy is associated to less accidents than Northern Italy, other variables being equal.

Table 6 – Alternative Italian NB model including the regional variable (p-values in brackets).

Model	**Parameters**						**AIC**
	β_0	**β_{AADT}**	**$\beta_{ST=1}$**	**$\beta_{ST=2}$**	**β_i**	**β_{REG}**	
IT(A)	-20.762 (<.001, 1.133)	1.397 (<.001, 0.118)	0.637 (<.001, 0.133)	0.982 (<.001, 0.163)	0.102 (0.001, 0.031)	-0.220 (0.096, 0.132)	1080.0

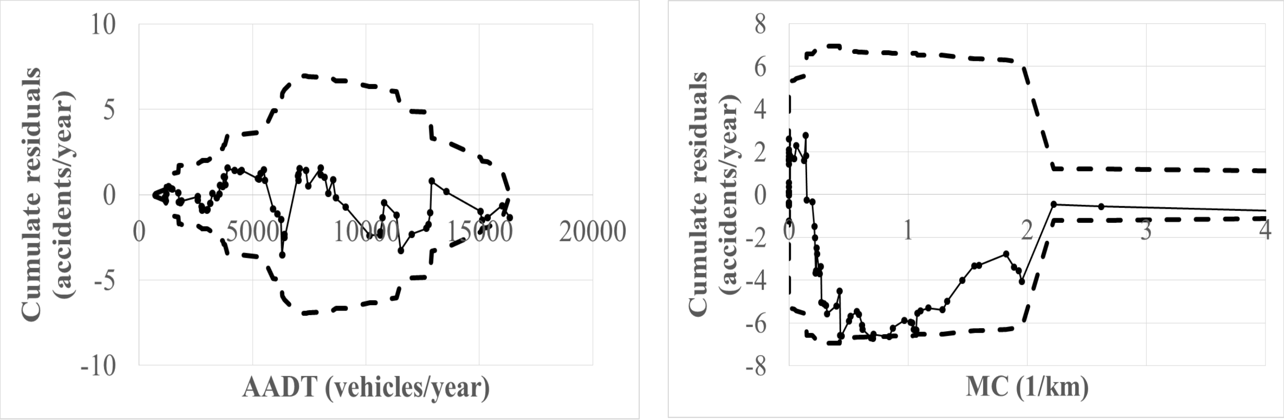

Figure 5. CURE plots for the Italian model (IT) related to the variables AADT and MC. Dashed lines represent the positive and negative two standard deviations (±2σ).

The analysis of the CURE plots in Fig. 5 reveals that the chosen model functional form is appropriate for the case of the AADT variable, with cumulate residuals oscillating around zero. Instead a significant overestimation effect of the model is revealed for the variable MC, in the range 0.2-0.6, and a subsequent underestimation effect in the range 0.6-2.0. In particular, this means that the chosen model significantly overestimate accident frequencies for low curvature elements, even if CURE are included in the confidence interval (dashed lines in Fig. 5), thus still implying acceptable results. The variable MC shows two high-leverage cases (MC > 4, truncated in Fig. 5 for graphical reasons). Note that a model similar to that shown in Table 5, estimated excluding these data, results in a slightly larger but comparable β_{MC} (about 0.4).

3.2 Scottish case study

3.2.1 Scottish HSM Calibration

Results from the HSM SPF calibration study for Scotland are reported as follows, including the assessment measure: $cv\{C_x\}$. They are further classified according to traffic and regions.

Table 7 – Results of the HSM SPF calibration study – Scotland

Variable: Region	**AADT Ranges**	**Number of Sites**	**Cx**	**cv[Cx]**	**Need for regional Cx (e_r) (Lord et al., 2016)**
Overall	**Overall**	311	0.71	0.12	-
	< 2,000	196	1.20	0.15	-
	≥ 2,000	115	0.48	0.17	-
“Lowlands” (South-West/East)	**Overall**	203	0.75	0.15	No (0.05)
	< 2,000	143	1.23	0.18	-
	≥ 2,000	60	0.41^	0.28	-
“Highlands” (Highlands-Islands/North-Eastern Scotland)	**Overall**	108	0.66	0.18	No (0.09)
	< 2,000	53	1.11^	0.30	-
	≥ 2,000	55	0.54^	0.21	-

Note: Cx coefficients marked with the superscript “^” are deemed less reliable due to either related number of segments < 30 or cv[Cx] ≥ 0.20. All subsets are associated to estimated “confidence levels” (based on Lord et al., 2016) significantly < 70 %.

Most calibration coefficients presented in Table 7 may be deemed reliable (Bahar and Hauer, 2014), except for the Highlands factors differentiated for traffic ranges and the Lowlands factor for traffic volumes < 2,000 (subsets having the smallest number of sites). The sample of sites considered (even if comparable with the Italian ones) lead to estimated “confidence levels” of calibration < 70 %, due to less observed accidents, for which SPF development would be preferable (Lord et al., 2016). A regional coefficient would not be needed for both the two regions considered.

From the analysis of data in Table 7, the HSM SPF generally overestimates accident frequencies for Scottish two-lane rural roads (the overall and most of the other Cx factors are < 1). There is a notable difference between traffic ranges: Cx are considerably higher for low traffic volumes (< 2,000) than high volumes. This result is valid for the overall estimate (i.e. Cx = 1.20 for low volume sites and Cx = 0.48 for high volume sites) and even disaggregating data regionally. Hence, the overestimation effect of the HSM SPF (Cx < 1) is amplified for traffic volumes > 2,000 (associated to low Cx values). The most reliable Cx factors showing low $cv\{C_x\}$ values are those obtained for the overall estimate and the first-level classification in regions and traffic ranges (i.e. not combining regions with traffic ranges). The Scottish calibration does not highlight any significant regional effect. The overall factor for the Lowlands (Cx = 0.75) is comparable to the Highlands (Cx = 0.66), as expected from the assessment procedure (no need for determining regional factors, based on Lord et al., 2016). This similarity can be noted even disaggregating according to the different traffic ranges.

3.2.2 Local Safety Performance Function: Scotland

The statistical parameters related to the fitted Scottish SPF are presented in Table 8, together with the over-dispersion parameters □. The model satisfactorily fits accident data, according to goodness-of-fit measures. However, the pseudo-R^2 is considerably lower than the Italian model, and the over-dispersion parameter is greater.

The variables included in the model are: shoulder type and weighted curvature. They are all significant at the chosen significance level (*p = 0.10*). Shoulders made of turf (ST = 1) are negatively related to the accident frequency (i.e. less accidents in presence of turf shoulders), with respect to paved/mixed shoulders (reference condition: ST = 0). Weighted curvature is positively related to the accident frequency, such as in the Italian case.

The following variables did not result statistically significant at the chosen significance level (*p = 0.10*): AADT, longitudinal slope, total road width, driveway density, RHR, region. Moreover, the analysis of the CURE plot in Fig. 6 reveals that the chosen model functional form is appropriate for what concerns the MC variable, with cumulate residuals oscillating around zero.

Table 8 – NB model parameters and goodness of fit measures for the Scottish SPF, with p-values in brackets

Model	**Parameters**			**Goodness-of-fit**			**Over-dispersion**
	β_0	$\beta_{ST=1}$	B_{MC}	**AIC**	χ^2	$_{Pseudo}R^2$	□
SC	-8.625 (<.001)	-0.399 (0.057)	0.122 (0.022)	360.2	199.5 (0.602)	0.171	6.630 (8.530)

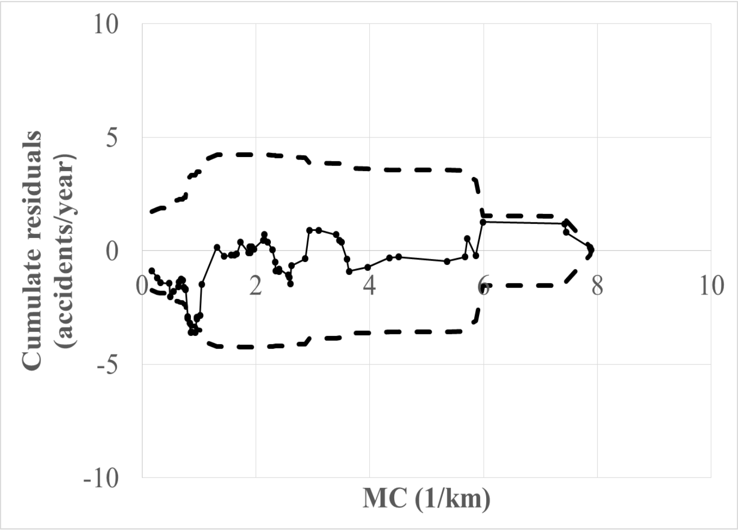

Figure 6. CURE plot for the Scottish model (SC) related to the variable MC. Dashed lines represent the positive and negative two standard deviations (±2σ).

4. DISCUSSION

Results obtained from both the TF calibration and LF estimation are discussed as follows, differentiated according to the main research questions posed in this study.

4.1 Calibration studies versus local SPF development

The first research question concerned the assessment of the general predictive capabilities of two different strategies (TF calibration and LF estimation), based on the case studies.

Calibration studies may be less demanding than LF estimations (especially if calibrations are conducted on base models including only some variables, e.g. traffic volumes, differently than the HSM calibration procedure, requiring several variables) and they may be conducted by non-experts through specific operational guidelines. However, the number of possible variables to take into account while conducting calibrations is some way limited by the necessary sample size for each combination of the considered variables. In fact, the reliability of a calibration factor may increase with the sample size, and minimum number of sites are suggested for calibration procedures (AASHTO, 2010; Lord et al., 2016). In this case, traffic, regions and the combinations of traffic ranges and regions were considered as detailed disaggregation of the TF calibration study (i.e. a calibration factor was derived for each combination of these variables). This means that several other categories may have been considered, by further disaggregating the sample in small samples (e.g. variables considered in the LF estimation: road width, curves, etc.).

Hence, considering only some variables for conducting detailed calibrations of TFs may lead to hide the influence of other variables. For example, while in the Italian case, a regional variability was noted, in the Scottish case, the LF development revealed other variables as influential on accident frequency (i.e. shoulder type and curve ratio) rather than geographic variables. Thus, a detailed Scottish calibration of TFs should include at least those other variables beyond regions, to ensure that the influence of geographic variables does not hide other strong relationships. However, as indicated above, this may imply an unbearable increase in the sample size (and information collected for each segment) for a simple calibration study. Moreover, the Scottish calibration proved to give unreliable indications about the role of traffic volume. Significant differences seemed to be present between low-volume (AADT < 2,000) and other segments. However, the variable AADT was not included in the Scottish model due to its lack of statistical significance. A zero-gradient relationship may actually exist between traffic and accidents, thus explaining the concurrent low calibration factor for high volumes and the high calibration factor for low volumes. This may be another argument for proceeding cautiously while selecting variables for calibration, even with variables usually associated with crashes (such as traffic volume).

On the other hand, several variables may be included in SPF modelling, being the mutual influence between predictors on the dependent variable considered as a part of the process. However, the data collection stage is more complex than a calibration study, due to the information required for each variable considered; and statistical applications are required. In LF estimation, some important variables may be excluded from best fitting models, due to their lack of statistical significance. However, on the contrary, disaggregating calibration factors according to different variables (e.g. traffic and region) and assessing their validity based on statistical indexes, may be misleading since the concurrent influence of other important variables may be ignored.

For what concerns the regional variability, TF calibrations may provide different calibration factors, but geographic variables may be excluded from finally selected models, as occurred in this study. Hence, calibration factors for TFs (even disaggregated according to different variables) should be carefully adopted. Their use may be justified in case of not available/obtainable LF. However, as noted in this present study, if a TF is calibrated, other road/traffic related variables should be preferred to regional variables, given the small dataset size.

For what concerns the general specific predictive capabilities of the calibrated TFs and estimated LFs in this study, they are assessed based on computed residuals (difference between observed and predicted values of yearly accident frequencies). To reveal possible significant improvements in the prediction, residuals were computed for each of the subsets considered for calibration (overall, regionally divided, classified into traffic ranges, classified into regions and traffic ranges). To allow the comparison between different calibrated TFs and estimated LFs, the synthetic measure: MAD (Mean Absolute Deviation) was used (such as in previous studies, see Oh et al., 2003; Sacchi et al., 2012; La Torre et al., 2014). It is obtained as the sum of the absolute residuals computed for each segment in the sample, divided by the number of segments. The closer the MAD index is to zero, the more the prediction is accurate. The obtained MADs are reported in the following Table 9.

Table 9 – Comparison of the Mean Absolute Deviation (MAD) [accidents/year] for the calibrated TFs and the estimated LFs in this study

Geographic area	Overall Calibration	Regional Calibration	Calibration with Traffic Ranges	Regional Calibration with Traffic Ranges	Local SPF
ITALY	0.623	0.590	0.585	0.581	0.541
SCOTLAND	0.430	0.433	0.386	0.384	0.365

An improvement in the prediction is noted for LFs with respect to calibrated HSM SPFs for both the Italian and Scottish case studies. An improvement is also noted if different regional and traffic subsets leading to specific calibration factors are considered, with respect to an overall calibration factor. As expected, the most relevant prediction improvement is noted while comparing MAD indexes of the locally developed SPF with the calibrated SPF. Paired t-tests were carried out to check the significance of the difference of the average MAD of corresponding calibrated and local SPFs. At the 5 % significance level no statistically significant difference was detected.

This further result has several implications in light of the aims of this study. In fact, it is important to note that even if the prediction capabilities of estimated LFs are greater than those of calibrated TFs (overall and disaggregated), the differences are not statistically significant. This means that the effort of developing a novel SPF, based on the same sample which can be used for HSM calibration, may be not justified by a significant prediction improvement. Even if this conclusion is solely based on the two case studies considered and the associated samples of road segments, it may have important practical consequence. In this sense, it should be also noted that the LFs developed in this study are based on small sample sizes and small sample means of observed accidents. This may lead to biased estimations, including unreliable over-dispersion parameters, which may severely influence the expected accidents resulting from the application of the Empirical Bayesian (EB) method (Lord, 2006). Hence, the development of local SPFs may be justified only in case of very large sample size, far greater than those required for HSM calibration, and in presence of several road and traffic variables collected. All these circumstances may lead to reliable and robust SPFs, which may significantly improve prediction capabilities with respect to simple calibrations. Otherwise, a detailed TF calibration (i.e. by at least considering the variability of traffic ranges) may represent a possible trade-off between computational, time and cost efforts and the reliability of results.

4.2 Influence of geographic variability on crash predictions

The second research question concerned the possible significance of geographic variables among all the variables used in predictive methods. In this study, the influence of geographic variability on crash predictions was explored through regional and terrain variables. The "region" variable (Italy: North, Centre-South; Scotland: "Highlands", "Lowlands") was considered in both the TF calibration and LF development. The "terrain" variable was considered in the Italian LF development.

The regional variability does not add significant explanations of the accident frequency as a result of the Scottish LF. In fact, region was not a significant variable included in the final model. Whereas, in the Italian study, while terrain was not a significant predictor, the region variable was included in a model alternative to the model associated with the lowest AIC measure. If the model in Table 6 would be used for accident prediction, estimates for Central-Southern Italy should be multiplied by $\exp(\boldsymbol{\beta}_{\mathbf{REG}})$, that is about 20 % smaller than predictions for Northern Italy, other conditions being equal. However, the final Italian model selected does not include the regional variable, but rather curvature and shoulder types, due to the associated improvement in the AIC score. The selection of model in Table 5 is not only due to merely computational considerations. In fact, while regional classifications may not be strongly influential on accident predictions, the influence of curvature is widely documented (see e.g. Abdel-Aty and Radwan, 2000; Elvik, 2013b). Thus, the model in Table 5 (including curvature but excluding regions) was definitely preferred.

A notable difference between calibration factors of Northern and Centre-Southern Italy (low traffic range: < 10,000) was noted, as expected from guidance by Lord et al. (2016). This may indicate that more crashes may be experienced in the Northern Italian low volume road segments, in respect to the Centre-Southern Italian corresponding segments. However, that Northern factor is deemed slightly unreliable. The same effect was noted in the intermediate SPF modelling stages (before selecting the final model), as discussed above. Thus, some influence of regional variability was revealed in the Italian case, from both the TF calibration and LF estimation. However, it should be noted that Northern Italian sites included in the sample are mostly high traffic volume sites (see Tables 1 and 5), differently from Central-Southern sites (mostly low-volume). SPF modelling should account for other variables (i.e. in this case traffic), while assessing the influence of a given variable (i.e. in this case region). However, it cannot be excluded that the significant difference in traffic volumes between the Northern and Central-Southern sites may hide the influence of other variables (not considered here) associated e.g. to the road importance, and which may have explained part of the variance, instead of a simple "region" variable. Hence, the regional variability issue for Italian accident predictions should be deepened in further studies with greater and homogeneous samples.

4.3 Geographic variability of accident predictors: Italy versus Scotland

The third research question concerned the possible discrepancies in the application of the considered predictive methods if different geographic contexts are considered. In this study, the two approaches (TF calibration and LF modelling) were repeated for both Italy and Scotland. Some macro differences between explanatory variables were highlighted indeed.

A remarkable difference between the two case studies is the role of traffic volume in explaining accidents. Traffic volume is often the most influential variable in predicting accident frequency. This is confirmed in the Italian study, but not in the Scottish one. The exclusion of the traffic variable from the Scottish SPF may seem surprising. However, the mean AADT for the Scottish sites is 2,048 vehi./day (st. dev.: 1,620 vehi./day); while the mean AADT for the Italian sites is 6,506 vehi./day (st. dev.: 4,269 vehi./day), thus having a wider spectrum of traffic volumes. The difference in the road networks of the two territories has contributed to the high discrepancy in traffic volumes. Secondary

Italian roads have mean traffic volumes greater than secondary Scottish roads, considering also that Scottish two-lane "A" class rural roads (likely with more traffic than secondary roads) were excluded from the database, because they belong to the primary network. However, it could be interesting to compare the accidents-traffic relationship for the same traffic volume interval (≤ 7,011, maximum Scottish volume). The cumulative frequencies of both accidents and traffic volumes are reported in Figure 7, considering both database, and the Italian database with comparable volumes (≤ 7,011).

As can be noted in Fig. 7, both Scotland and Italy (for the same low-volume traffic range: ≤ 7,011 vehi./day) exhibit a relevant frequency of zero-count sites (30-40 %), and a similar distribution of the cumulative frequency of accidents/km (Fig. 7). However, when it comes to traffic volumes, there is a notable difference between Italy and Scotland. Scottish volumes are heavily skewed on a very-low volume (i.e. approx.. 40 % of sites have AADT ≤ 1,500, and 75 % of sites have AADT ≤ 3,000), while Italian sites are not. This may have affected the search for a satisfactorily fitting accidents-traffic curve. To note, an attempt Italian model fitted by considering only sites having AADT ≤ 7,011 still revealed traffic volume as a significant variable, even if with β_{AADT} close to 1, instead of > 1.

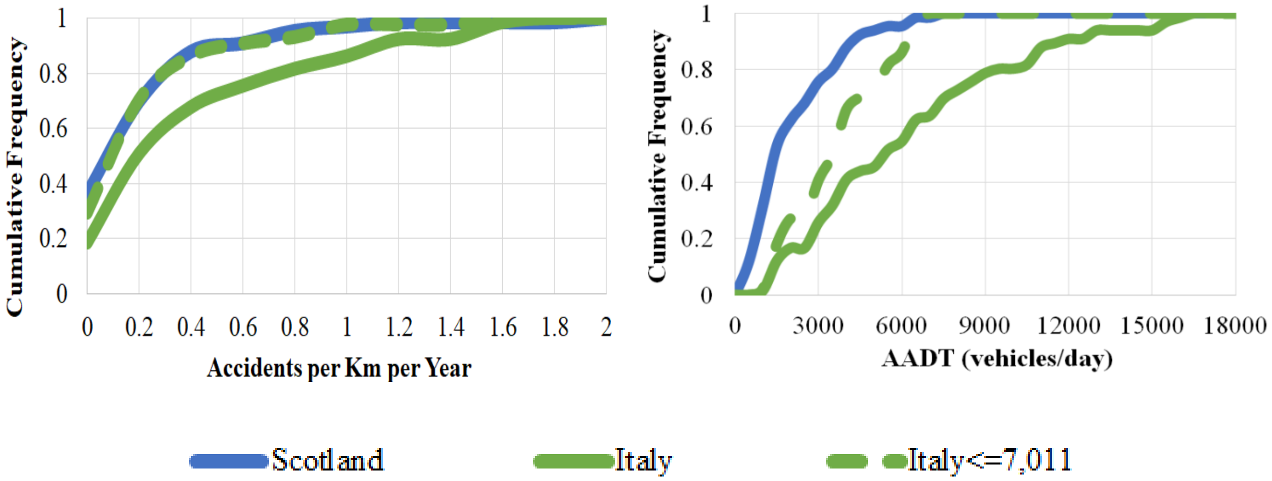

Figure 7 – Cumulative fatal+injury accident frequencies and traffic volumes for Italy and Scotland

The above reported findings lead to the following remarks, which are of practical interest for researchers and, to some extent, for road safety practitioners:

- In case of a sample of secondary two-lane road sites having low traffic volumes and also skewed to very-low volumes, the accident frequency may not significantly be dependent on the amount of traffic volumes (as found for Scotland). This could be explained by the very low number of interactions between vehicles in the traffic flow, and most of the accidents may be single-vehicle accidents (e.g. run-off-road). This should be confirmed by future studies conducted on sites with AADT similar to the Scottish sites. Moreover, in this case, as explained above, different calibration factors obtained for different traffic ranges (as in this case, using 2,000 as a threshold) may be unreliable even if statistically valid (Table 4).
- In case of a sample of two-lane road sites having a wide spectrum of traffic volumes as the Italian ones, the relationship between accident and traffic was found to be more than linear ($\beta_{AADT} \sim 1.4$). However, when separating only sites with AADT ≤ 7,011 (comparable to the Scottish ones), the inferred accidents-traffic relationship becomes approximately linear. Hence, a linear relationship as the one in the HSM (AASHTO, 2010) may only be valid for low traffic volumes (approximately < 10,000, see Sacchi et al., 2012). Hence, in case of sites with widely varying traffic volumes, different traffic ranges should be considered if only

calibration is conducted. In this way, the nature of the non-linear accidents-traffic relationship may be captured also in a calibration procedure (see Table 4, $Cx_{,\geq 10,000} >> Cx_{,<10,000}$).

The effect of curvature is strongly related to accident frequencies, as found in previous studies (e.g. Abdel-Aty and Radwan, 2000; Elvik, 2013b). This is valid for both Italy and Scotland. The effect of curvature is more evident on Italian than on Scottish sites, by comparing the β_{MC} coefficients. This may be explained by the nature of road sites considered. Italian sites have mean CR: 0.14 (st. dev.: 0.15), mean radius of curvature: 295 m (st. dev.: 195 m), while Scottish sites have mean CR: 0.55 (st. dev.: 0.26), mean radius of curvature: 349 m (st. dev.: 275 m). Hence, mean radii of curvature of curved segments are similar, while the percentages of curved sites on the segment (CR) are not. Scottish segments are notably more winding than the Italian ones. The small segment curvature may lead Scottish drivers to select lower speeds and this, in turn, may result in lower accident risks (Aarts and Van Schagen, 2006; Elvik, 2013a). The reduced accident risk may also be due to the smaller skidding risk at low speeds (Colonna et al., 2016b). On the other hand, Italian drivers may select higher speeds on the sample of road sites due to the low percentage of curves. Because of the higher Scottish segment curve ratio, the mean speed differential between consecutive segments and curves (especially if sharp) for Scottish drivers may likely be lower than the corresponding Italian drivers' speed differential. The inclusion of variables which attempt at capturing operating speeds and speed differences (see e.g. Cafiso et al., 2010) may have helped in revealing those differences related to samples of roads with different importance. Since it was not possible to derive those variables from the dataset inquired, further research on the regional variability of accident predictions should consider also speed variables. Local operating speed models (see e.g. Discetti et al., 2011) may help for this aim, even if relying on a predicted operating speed as a base variable for SPFs may lead to an increase in both the uncertainty and the unreliability of results.

The effect of different shoulder types (paved, unpaved, mixed/composite) is related as well to accident frequencies, as expected from previous studies (see Zeeger and Deacon, 1987). However, the effect is different in the two case studies considered. In the Italian case, paved shoulders are the safest condition with respect to accident frequencies, while turf and composite/mixed/gravel shoulders are the less safe. This is in line with expectations from HSM (AASHTO, 2010). On the contrary, in the Scottish case, turf shoulders result as safer than mixed/composite and paved shoulders (to note, there is only one segment having paved shoulders). This difference may be explained again by the diverse importance of the road segment classes (low-volume Scottish and medium-volume Italian secondary roads). Roads with turf shoulders (the majority of Scottish sites: 62 %, largely different than Italian sites: only 22 %) may be an indirect indicator of the minor road importance, which can be travelled at relatively lower speeds. On the other hand, the presence of turf shoulders itself (as the case of narrow shoulders or reduced clearance, see e.g. Martens et al., 1997) may lead drivers to decrease their speeds, and then to better performances in terms of accident frequencies. However, the other category is mostly composed of unpaved shoulders as well, thus being the comparison with paved shoulders unfeasible in this case.

5. CONCLUSIONS

The issue of geographic variability of SPFs and associated predictors, both at the trans-national and the inner scales poses important questions to both researchers and road safety practitioners. Two European case studies (one for the Italian, the other for Scottish road sites) were analysed to provide new insights in this field, by using two different approaches: calibration of a transferred function (TF) or estimation of a local function (LF). The following conclusions are drawn, based on the results obtained from the two case studies, and their comparison:

- A trans-national variability of accident predictions was noted between Italy and Scotland. This was largely associated to the different nature of the two two-lane road networks. The representative Scottish road sites present lower traffic volumes and design features (i.e. more curves, unpaved shoulders, etc.) than the Italian sample of sites. This affected the modelling stage, revealing a not significant influence of traffic on Scottish accidents. The highlighted result and the possible existence of traffic volume thresholds below which the influence of traffic decreases should be verified in future studies for very-low volume roads.
- An inner variability of accident predictions was not found in the Scottish case, while it was individuated in the Italian case study (in both calibration and the intermediate stages of SPF development and selection). However, as explained in the text, a weak regional variability may rather hide the influence of other variables. Anyway, the finally selected Italian model did not include region as a significant predictor. This may lead to conclude that time and costs necessary for considering geographic variability of crash predictions among administrative boundaries may be saved, by prioritizing other variables. The homogeneity of road design standards among countries may be prevalent on local differences (e.g. drivers' behaviour). This was evident in Scotland, while further studies could be needed in the Italian case.
- Calibration procedures (especially those accounting only for some variables) may be inexpensive and easier than LF estimation. However, even statistically significant calibration factors may be "false positives" when checked against results of a comprehensive SPF, such as the differences between traffic ranges and regions in this study. On the other hand, LF estimations based on the same sample size required for TF calibrations may only slightly improve the predictive capabilities of a simple TF calibration, as revealed in this study. Hence, when sufficiently large and statistically representative sample size, and the related detailed datasets of road/traffic features are not available, the efforts for estimating a new LF could be saved and the TF calibration could be a good compromise (e.g. for practitioners, when LFs are not available).
- The segment curvature and the shoulder types were revealed as significant crash predictors in both the Italian and Scottish models, even with some local differences, attributed to the different importance of roads and their possible influence on speeds (which were not modelled in this study). Road width, elevation, roadside hazard, driveway density and longitudinal slopes resulted not statistically significant accident predictors in both models.

Clearly, those conclusions are based on the two analysed case studies and the associated database. As explained in the text, due to the wide variability of all the factors involved in the accident predictions, these results may be neither generalized to a wider scale, nor applicable in other different jurisdictions. This is also the main limitation of this study, which is intrinsic of SPF development and calibration procedures. To note, greater samples of sites may have potentially improved the model fit or the significance of calibration coefficients, allowing more combinations of variables. However, due to several layers of analyses conducted in a single study, the database considered were deemed satisfactorily representative. Moreover, the two presented models for Italy and Scotland, represent an immediate applicable tool for road safety practitioners, especially for the Scottish secondary road network, for which no previous similar studies were found. However, in the Scottish case, further research is needed to provide new insights about traffic volume-accidents relationships on very low-volume roads.

Given the importance of the topic for road planning and design purposes and the need for guidance to select the best predictive approach in each local area, future research should be focused in improving and enlarging the knowledge in this field. This means that assessments similar to those performed in this article should be ideally conducted for each country/state. At the local level, future research should confirm the weak importance of regional and terrain characteristics in the considered contexts, especially in the Scottish case.

ACKNOWLEDGEMENTS

Rita Binetti wishes to acknowledge the European Union (Erasmus Traineeship Program) and the Polytechnic University of Bari for providing support for her research at the Edinburgh Napier University.

REFERENCES

Aarts, L., & Van Schagen, I. (2006). Driving speed and the risk of road crashes: A review. *Accident Analysis & Prevention*, 38(2), 215-224.

Abdel-Aty, M. A., & Radwan, A. E. (2000). Modeling traffic accident occurrence and involvement. *Accident Analysis & Prevention*, 32(5), 633-642.

American Association of State Highway and Transportation Officials (AASHTO) (2010). *Highway Safety Manual*. Washington, DC.

Bahar, G. B., & Hauer, E. NCHRP 20-07 (332) (2014) *Users' Guide to Develop Highway Safety Manual Safety Performance Function Calibration Factors*, Transportation Research Board, Washington, D.C.

Bauer, K. M., & Harwood, D. W. (2000). *Statistical Models of At-Grade Intersection Accidents*. Addendum (Report No. FHWA-RD-99-094). United States Federal Highway Administration.

Brimley, B., Saito, M., & Schultz, G. (2012). Calibration of Highway Safety Manual safety performance function: development of new models for rural two-lane two-way highways. *Transportation Research Record: Journal of the Transportation Research Board*, (2279), 82-89.

Cafiso, S., Di Graziano, A., Di Silvestro, G., La Cava, G., & Persaud, B. (2010). Development of comprehensive accident models for two-lane rural highways using exposure, geometry, consistency and context variables. *Accident Analysis & Prevention*, 42(4), 1072-1079.

Cafiso, S., Di Silvestro, G., & Di Guardo, G. (2012) Application of Highway Safety Manual to Italian divided multilane highways. *Procedia-Social and Behavioral Sciences*, 53, 911-920.

Chatterjee, S., & Simonoff, J. S. (2013). *Handbook of regression analysis*. John Wiley & Sons.

Colonna, P., Berloco, N., Intini, P., Perruccio, A., Ranieri, V., & Vitucci, V. (2016a). Variability of the Calibration Factors of The HSM Safety Performance Functions with Traffic, Region and Terrain. The case of the Italian rural two-lane undivided road network. *Compendium of Papers of the 95th Annual Meeting of the Transportation Research Board* (No. 16-3413).

Colonna, P., Berloco, N., Intini, P., Perruccio, A., & Ranieri, V. (2016b). Evaluating skidding risk of a road layout for all types of vehicles. *Transportation Research Record: Journal of the Transportation Research Board*, (2591), 94-102.

Colonna, P., Intini, P., Berloco, N., & Ranieri, V. (2018). Integrated American-European protocol for safety interventions on existing two-lane rural roads. *European transport research review*, 10(1), 5.

Davidson, N. (2000). *The origins of Scottish nationhood*. Pluto Press.

Devine, T. M. (1979). Temporary migration and the Scottish Highlands in the nineteenth century. *The Economic History Review*, 32(3), 344-359.

Discetti, P., Dell'Acqua, G., & Lamberti, R. (2011). Models of operating speeds for low-volume roads. *Transportation Research Record: Journal of the Transportation Research Board*, 2203(1), 219-225.

Donnell, E., Gayah, V. & Jovanis, P. (2014). *Safety Performance Functions*. Pennsylvania Department of Transportation. Report no. FHWA-PA-2014-007-PSU WO 1.

Elvik, R. (2013a). A re-parameterisation of the Power Model of the relationship between the speed of traffic and the number of accidents and accident victims. *Accident Analysis & Prevention*, 50, 854-860.

Elvik, R. (2013b). International transferability of accident modification functions for horizontal curves. *Accident Analysis & Prevention*, 59, 487-496.

European Parliament and Council (2003). *Regulation (EC) No 1059/2003 of the European Parliament and of the Council of 26 May 2003 on the establishment of a common classification of territorial units for statistics (NUTS).*

European Union. Eurostat. http://ec.europa.eu/eurostat/web/nuts/nuts-maps-.pdf-. Last accessed on June, 6th, 2018. http://ec.europa.eu/eurostat/data/database. Last accessed on August, 8th, 2018.

European Union (2015). Regions in the European Union. Nomenclature of territorial units for statistics. NUTS 2013/EU-28. Eurostat. Manuals and Guidelines.

Farid, A., Abdel-Aty, M., Lee, J., Eluru, N., & Wang, J. H. (2016). Exploring the transferability of safety performance functions. *Accident Analysis & Prevention*, 94, 143-152.

Farid, A., Abdel-Aty, M., & Lee, J. (2018a). A new approach for calibrating safety performance functions. *Accident Analysis & Prevention*, 119, 188-194.

Farid, A., Abdel-Aty, M., & Lee, J. (2018b). Transferring and calibrating safety performance functions among multiple states. *Accident Analysis & Prevention*, 117, 276-287.

Farid, A., Abdel-Aty, M., & Lee, J. (2019). Comparative analysis of multiple techniques for developing and transferring safety performance functions. *Accident Analysis & Prevention*, 122, 85-98.

Garach, L., de Oña, J., López, G., & Baena, L. (2016). Development of safety performance functions for Spanish two-lane rural highways on flat terrain. *Accident Analysis & Prevention*, 95, 250-265.

Garber, N. J., Haas, P. R., & Gosse, C. (2010). *Development of safety performance functions for two-lane roads maintained by the Virginia Department of Transportation* (No. FHWA/VTRC 10-R25).

Geedipally, S.R., Shirazi, M., Lord, D. (2017). *Exploring the need for region-specific calibration factors*. Transportation Research Record, 2636, 73–79.

Gomes, S. V., Geedipally, S. R., & Lord, D. (2012). Estimating the safety performance of urban intersections in Lisbon, Portugal. *Safety science*, 50(9), 1732-1739.

Gomes, M. M., Pirdavani, A., Brijs, T., & Pitombo, C. S. (2019). Assessing the impacts of enriched information on crash prediction performance. *Accident Analysis & Prevention*, 122, 162-171.

González, M. P. S., Sotos, F. E., & Ponce, Á. T. (2018). Impact of provincial characteristics on the number of traffic accident victims on interurban roads in Spain. *Accident Analysis & Prevention*.

Gooch, J. P., Gayah, V. V., & Donnell, E. T. (2018). Safety performance functions for horizontal curves and tangents on two lane, two way rural roads. *Accident Analysis & Prevention*, 120, 28-37.

Greibe, P. (2003). Accident prediction models for urban roads. *Accident Analysis & Prevention*, 35(2), 273-285.

Hauer, E. (1999). *Safety in geometric design standards*. University of Toronto, Department of Civil Engineering.

Hauer, E., & Bamfo, J. (1997). Two tools for finding what function links the dependent variable to the explanatory variables. In *Proceedings of the ICTCT 1997 Conference*, Lund, Sweden.

Hauer, E., Bonneson, J., Council, F., Srinivasan, R., Zegeer, C. (2012). Crash modification factors: foundational issues. *Transportation Research Record: Journal of the Transportation Research Board*, 2279, 67–74.

Hauer, E., Persaud, B. (1997). *Safety analysis of roadway geometric and ancillary features.* Research Report. Transportation Association of Canada.

Hauer, E. (2015). *The art of regression modeling in road safety* (Vol. 38). New York: Springer.

Hilbe, J. M. (2011). *Negative binomial regression*. Cambridge University Press.

ISTAT. Istituto Nazionale di Statistica. [Italian National Institute of Statistics]. http://dati.istat.it/. Last accessed on June, 05th, 2018.

Italian Ministry of Infrastructures and Transport (2001). Functional and Geometric standards for road design. D.M. 6792.

Jackman, S. (2017). Political Science Computational Laboratory, ‘pscl’ package v.1.5.2.

Kononov, J., & Allery, B. (2003). Level of service of safety: Conceptual blueprint and analytical framework. *Transportation Research Record: Journal of the Transportation Research Board*, 1840, 57-66.

La Torre, F., Domenichini, L., Corsi, F., & Fanfani, F. (2014). Transferability of the Highway Safety Manual Freeway Model to the Italian Motorway Network. *Transportation Research Record: Journal of the Transportation Research Board*, 2435, 61-71.

Lee, J., Abdel-Aty, M., & Cai, Q. (2017). Intersection crash prediction modeling with macro-level data from various geographic units. *Accident Analysis & Prevention*, 102, 213-226.

Li, L., Gayah, V. V., & Donnell, E. T. (2017). Development of regionalized SPFs for two-lane rural roads in Pennsylvania. *Accident Analysis & Prevention*, 108, 343-353.

Liu, J., Khattak, A. J., & Wali, B. (2017). Do safety performance functions used for predicting crash frequency vary across space? Applying geographically weighted regressions to account for spatial heterogeneity. *Accident Analysis & Prevention*, 109, 132-142.

Lord, D., & Mannering, F. (2010). The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives. *Transportation Research Part A: Policy and Practice*, 44(5), 291-305.

Lord, D., Manar, A., & Vizioli, A. (2005a). Modeling crash-flow-density and crash-flow-V/C ratio relationships for rural and urban freeway segments. *Accident Analysis & Prevention*, 37(1), 185-199.

Lord, D., Washington, S. P., & Ivan, J. N. (2005b). Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory. *Accident Analysis & Prevention*, 37(1), 35-46.

Lord, D. (2006). Modeling motor vehicle crashes using Poisson-gamma models: Examining the effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter. *Accident Analysis & Prevention*, 38 (4), 751-766.

Lord, D., Geedipally, S. R., & Shirazi, M. (2016). *Improved guidelines for estimating the Highway safety manual calibration factors* (No. ATLAS-2015-10). University Transportation Centers Program (US).

Martens, M. H., Compte, S., & Kaptein, N. A. (1997). *The effects of road design on speed behaviour: a literature review*. Deliverable D1 (Report 2.3.1), MASTER.

McCullagh, P. (1984). Generalized linear models. *European Journal of Operational Research*, 16(3), 285-292.

Myers, R. H., Montgomery, D. C., Vining, G. G., & Robinson, T. J. (2012). *Generalized linear models: with applications in engineering and the sciences*. John Wiley & Sons.

Oh, J., Lyon, C., Washington, S., Persaud, B., & Bared, J. (2003). Validation of FHWA crash models for rural intersections: Lessons learned. *Transportation Research 26 Record: Journal of the Transportation Research Board*, (1840), pp. 41-49.

Oh, J., Washington, S.P., Nam, D. (2006). Accident prediction model for railway highway interfaces. *Accident Analysis and Prevention* 38 (2), 346–356.

Persaud, B., Lord, D., & Palmisano, J. (2002). Calibration and transferability of accident prediction models for urban intersections. *Transportation Research Record: Journal of the Transportation Research Board*, (1784), 57-64.

PRACT Project. Predicting Road ACcidents - a Transferable methodology across Europe. http://www.practproject.eu/. Last accessed on June, 05th, 2018.

Qin, X., Ivan, J. N., Ravishanker, N., & Liu, J. (2005). Hierarchical Bayesian Estimation of Safety Performance Functions for Two-lane Highways using Markov Chain Monte Carlo modeling. *Journal of Transportation Engineering,* 131(5): 345-351.

Russo, F., Busiello, M., & Dell, G. (2016). Safety performance functions for crash severity on undivided rural roads. *Accident Analysis & Prevention*, 93, 75-91.

Sacchi, E., Persaud, B., & Bassani, M. (2012). Assessing international transferability of highway safety manual crash prediction algorithm and its components. *Transportation Research Record: Journal of the Transportation Research Board*, 2279, 90-98.

Sawalha, Z., & Sayed, T. (2006). Transferability of accident prediction models. *Safety science*, 44(3), 209-219.

Scottish Office for National Statistics. Detailed information on the administrative structure within Scotland.https://www.ons.gov.uk/methodology/geography/ukgeographies/administrativegeography/scotland. Last accessed on June, 05th, 2018.

Shin, H. S., Dadvar, S., & Lee, Y. J. (2015) Results and Lessons from the Local Calibration Process of the Highway Safety Manual for the State of Maryland. In *Transportation Research Board 94th Annual Meeting* (No. 15-4643).

Shirazi, M., Geedipally, S. R., Lord, D. (2016)a. A procedure to determine when safety performance functions should be recalibrated. *Journal of Transportation Safety & Security,* 9(4), 459-467.

Shirazi, M., Lord, D., Geedipally, S. R. (2016)b. Sample-size guidelines for recalibrating crash prediction models: Recommendations for the Highway Safety Manual. *Accident Analysis & Prevention*, 93, 160-168.

Srinivasan, R., & Carter, D. L. (2011). *Development of safety performance functions for North Carolina.* Chapel Hill, NC: North Carolina Department of Transportation, Research and Analysis Group. (Report No. FHWA/NC/2010-09).

Srinivasan, R., Colety, M., Bahar, G., Crowther, B., & Farmen, M. (2016). Estimation of calibration functions for predicting crashes on rural two-lane roads in Arizona. *Transportation Research Record: Journal of the Transportation Research Board*, (2583), 17-24.

Sun, X., Li, Y., Magri, D., & Shirazi, H. (2006). Application of highway safety manual draft chapter: Louisiana experience. *Transportation Research Record: Journal of the Transportation Research Board*, 1950, 55-64.

Tarko, A. P. (2006). Calibration of safety prediction models for planning transportation networks. *Transportation Research Record: Journal of the Transportation Research Board*, 1950(1), 83-91.

Transport Scotland (2016). Scottish Transport Statistics. A National Statistics Publication for Scotland. 2016 Edition. No. 35.

UK Department for Transport (2012). Guidance on road classification and the primary route network.

Venables, W. N. & Ripley, B. D. (2002) Modern Applied Statistics with S. Fourth Edition. Springer, New York. ISBN 0-387-95457-0.

Washington, S. P., Karlaftis, M. G., & Mannering, F. (2010). Statistical and econometric methods for transportation data analysis. Chapman and Hall/CRC.

Xie, F., Gladhill, K., Dixon, K., & Monsere, C. (2011). Calibration of Highway Safety Manual Predictive Models for Oregon State Highways. *Transportation Research Record: Journal of the Transportation Research Board*, 2241, 19-28.

Yannis, G., Dragomanovits, A., Laiou, A., Richter, T., Ruhl, S., La Torre, F, Domenichini, L., Graham, D., Karathodorou, N. & Li, H. (2016). Use of accident prediction models in road safety management - an international inquiry. *Transportation Research Procedia*, 14, 4257-4266.

Zegeer, C. V., & Deacon, J. A. (1987). *Effect of lane width, shoulder width, and shoulder type on highway safety*. State of the art report, 6, 1-21.

Zhang, C., & Ivan, J. (2005). Effects of geometric characteristics on head-on crash incidence on two-lane roads in Connecticut. *Transportation Research Record: Journal of the Transportation Research Board*, 1908, 159-164.