



Politecnico di Bari

Repository Istituzionale dei Prodotti della Ricerca del Politecnico di Bari

Transferred versus local Safety Performance Functions: A geographical analysis considering two European case studies

This is a pre-print of the following article

Original Citation:

Transferred versus local Safety Performance Functions: A geographical analysis considering two European case studies / Intini, Paolo; Berloco, Nicola; Binetti, Rita; Fonzone, Achille; Ranieri, Vittorio; Colonna, Pasquale. - In: SAFETY SCIENCE. - ISSN 0925-7535. - STAMPA. - 120:(2019), pp. 906-921. [10.1016/j.ssci.2019.08.013]

Availability:

This version is available at <http://hdl.handle.net/11589/191757> since: 2021-03-03

Published version

DOI:10.1016/j.ssci.2019.08.013

Publisher:

Terms of use:

(Article begins on next page)

1 **TRANSFERRED VERSUS LOCAL SAFETY PERFORMANCE FUNCTIONS: A**
2 **GEOGRAPHICAL ANALYSIS CONSIDERING TWO EUROPEAN CASE STUDIES**

3 by **Paolo Intini, Nicola Berloco, Rita Binetti, Achille Fonzone, Vittorio Ranieri, Pasquale**
4 **Colonna**

5 **ABSTRACT**

6 Two main approaches can be used to predict road accidents: transferring existing Safety Performance
7 Functions (SPFs) from other areas (transferred SPFs) and developing local SPFs. Both approaches
8 have advantages and disadvantages and are affected by the difficult choice of predictors. Regional
9 variables or terrain factors may lead prediction improvements. However, results from previous
10 relevant research are contradictory and transferability assessments are mainly based on North-
11 American experiences. Because of these inconsistencies, this study is an attempt of providing new
12 insights on the choice between alternative accident prediction methods by taking into account the
13 geographic variability in the European context. In particular, it addresses three main issues: (1) it
14 compares the prediction accuracy of transferred and local SPFs; (2) it determines the significance of
15 regional factors in explaining safety performances, (3) it assesses the variability of results among the
16 different contexts considered. Research questions are addressed as based on two-lane rural road sites
17 in Italy and Scotland. The analysis shows differences between the two countries, due to the different
18 nature of the networks, but not within each country. Both advantages and disadvantages were
19 highlighted in the evaluation of transferred and local SPFs. Calibration of transferred SPFs may be
20 less demanding than their local estimation, even if they may lead to unreliable estimates when
21 compared to comprehensive SPFs. However, locally developed SPFs may not provide more
22 significantly reliable estimates than transferred SPFs. Segment curvature and shoulder types are
23 statistically significant predictors in both the Italian and Scottish models, even having different
24 importance.

25

26

1. INTRODUCTION

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

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

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

44 1.1 Background on transferability of accident predictive methods

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

46 The first strategy consists in transferring SPFs from other areas (Transferred Functions, TFs), and
47 calibrating them by correcting their outcomes according to local conditions. A possible basic
48 calibration method is provided in the Highway Safety Manual (HSM) (AASHTO, 2010). Local
49 calibration factors are computed as the ratio of the crashes observed on a sample of local road sites;
50 to those predicted by the base model for the same types of sites. However, a single calibration factor
51 could not be sufficient for large/not homogeneous areas (e.g. different terrains) (Bahar and Hauer,
52 2014). Hence, different calibration factors may be achieved in case of different local characteristics
53 (see e.g. Tarko, 2006; Colonna et al., 2016a). More refined calibration techniques were defined, which
54 may provide more reliable estimates. For example, the calibration of model parameters through
55 maximum likelihood estimation (Sawalha and Sayed, 2006); segment-specific calibration (Farid et
56 al., 2016); calibration functions (Srinivasan et al., 2016); calibration based on local regression (Farid
57 et al., 2018b) or on the k nearest neighbour data mining method (Farid et al., 2018a), were proposed.

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

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

78 **1.2 Background on the geographic variability of the transferability issues**

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

83 For example, concerning the first point, calibration factors for TFs may greatly vary for different
84 regions of the same country (Colonna et al., 2016a), or even in sub-networks of the same state (Tarko,
85 2006). However, country-wide calibrations were conducted as well (see e.g. La Torre et al., 2014).

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

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

110 Hence, it is evident how geographic factors (not only jurisdiction-related) may both affect the
111 transferability of SPFs and the development of calibration factors. Recent studies have then focused

112 on considering geographic factors for crash analyses at different levels: i.e. at the provincial level
113 while taking into account macro-variables (Gonzalez et al., 2018), or even more disaggregate levels
114 while considering a mix of macro and local variables (Lee et al., 2017). However, several variables
115 related to road geometry, traffic operations, and boundary conditions should be considered in the SPF
116 estimation (see e.g. Hauer, 2015). Given their consistent importance revealed in previous research
117 (e.g. Abdel-Aty and Radwan, 2000; Greibe, 2003; Cafiso et al., 2010), the assessment of geographic
118 variability should not be conducted independently from other road geometric and traffic variables.

119 **1.3 Research questions**

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

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

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

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

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

142

143 **2. METHODS**

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

146 **2.1 Procedure**

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

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

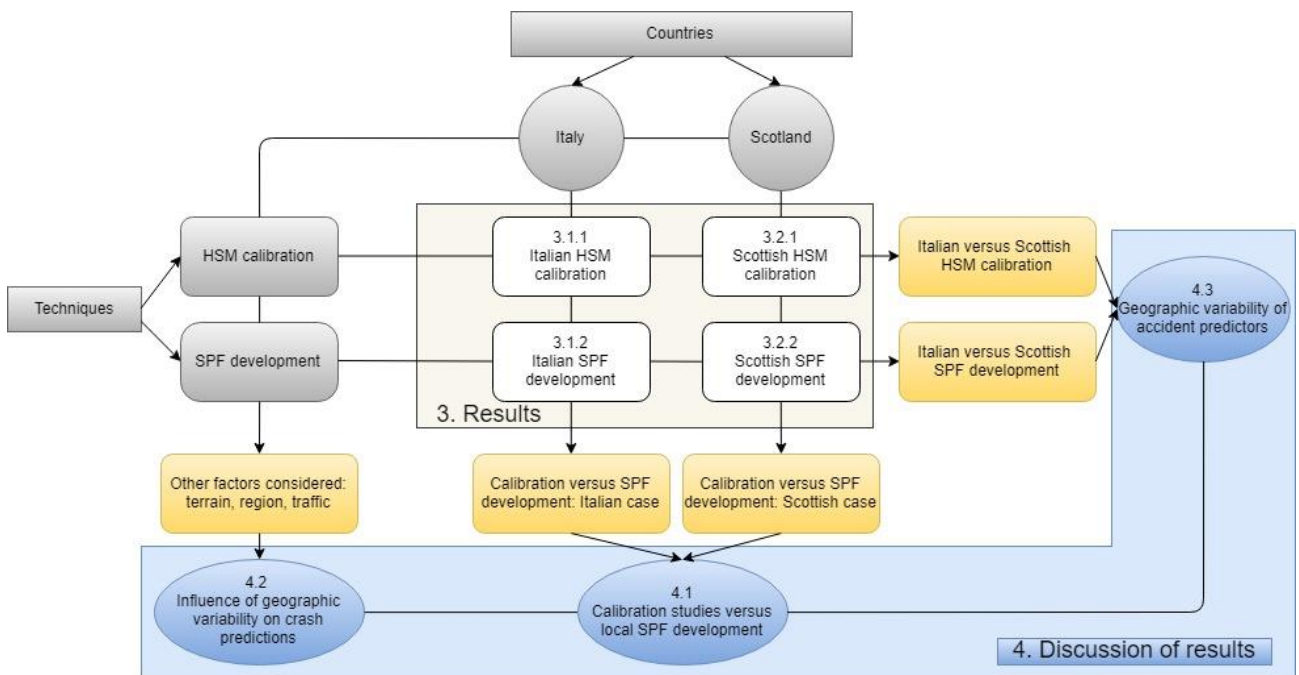
- 152 • Assess the general influence of geographic variability factors on crash predictions; i.e. if the
153 geographic factors (both different geographic areas and terrains) may influence the calibration
154 factors or be included in the regression analysis;
- 155 • Compare the results obtained through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

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

160 Different SPFs may have been considered for TF calibration for both Italy and Scotland. However,
161 the sequential application of HSM SPF and CMFs for two-lane rural roads includes a wider list of
162 road and traffic accident predictors than several other alternative models. For example, Colonna et al.
163 (2018) highlighted that the two-lane rural HSM SPF calibrated for Italy can account for several road
164 and traffic features, when compared with alternative Italian models (Cafiso et al., 2010; Russo et al.,
165 2016). Thus, the base HSM SPF (and related CMFs) were selected as they may represent a wide range
166 of road and traffic characteristics. Moreover, the HSM SPF represents a usual benchmark TF in
167 previous research (see e.g. Sacchi et al., 2012).

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

176



177

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

179 **2.2 Database**

180 Two separate databases, namely, for Italy and Scotland, were used. Both database are composed of a
 181 first traffic volume dataset, and a second accident (fatal+injury only) dataset for two-lane rural roads.
 182 Hence, only the secondary road networks of the two areas were considered, thus excluding roads
 183 belonging to the primary and main road networks (“A” and “B” class in the Italian classification,
 184 Italian Ministry of Infrastructures and Transport, 2001; motorways and “A” class in the UK
 185 classification, UK Department for Transport, 2012). Italian primary and main roads should be
 186 designed as multi-lane roads (whether being motorways or not). Whereas, the main UK roads (“A”
 187 class) may include also some two-lane roads. However, “A” class roads were not considered in the
 188 road network composed of secondary roads, to be coherent with the Italian case.
 189

190 Annual average daily traffic counts were collected from the respective road agencies (UK Department
 191 for Transport, covering all the Scottish network; Italian ANAS, covering part of the Italian network).
 192 Accident data were retrieved from different sources: Italian National Institute of Statistics (ISTAT)
 193 and Italian Automobile Club (ACI) for the Italian case and the online portal <https://data.gov.uk/> for
 194 the Scottish case. At least three years of accident data were collected (see Bahar and Hauer, 2014).
 195

196 Starting from the overall database, traffic and accident data were coupled for road sections provided
 197 with traffic counts. A road section is defined here as a section on a road trunk included between two
 198 relevant intersections (i.e. with roads of similar importance, excluding driveways or intersections with
 199 minor roads), on which a unique traffic volume is assigned, since it is deemed as constant along it.
 200 The resulting total length of segments inquired is about 213 km (74 segments) for Italy and 180 km
 201 for Scotland (66 segments).
 202

203 The total number of observed Scottish crashes is low (101 in total), even if the total length of segments
 204 investigated is comparable with the Italian one. Hence, among all the segments provided with traffic
 205 data, a subset was selected in compliance with both the following requirements: 1) having an
 206 equivalent number of at least 100 accidents/year (AASHTO, 2010), 2) including a sufficient number
 207 of zero-count sites to account for the low mean estimated accidents/km rate in the part of network
 208 investigated. Detailed information concerning the road segments composing the final database
 209 obtained are reported in the following Table. Information about the dataset are also classified
 210 according to the traffic ranges and regions of the segments, which pertains to the main research
 211 questions. Descriptive statistics are also reported about accidents, traffic, geometric and other
 212 characteristics of the segments in the dataset. The variables considered in this study are described in
 213 detail in 2.3.
 214

215 *Table 1. Descriptive statistics of the variables considered among the sample of segments, showing*
 216 *the mean values (st. dev. in brackets) or counts associated to each variable over the considered road*
 217 *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)

218

219 2.3 Variables

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

222 2.3.1 Geographic variables

223 Coherently with the study aims, geographic factors were considered within each country and not only
224 as the difference between countries (i.e. Italy versus Scotland). Hence, both Italy and Scotland were
225 divided into regions, used as synthetic variables to capture the influence of socio-economic and
226 driving behavioural differences. Italy (I) and Scotland (S) are hardly comparable in terms of area
227 (approx. 300,000 km² (I) and 80,000 km² (S)), population (approx. 60 million inhabitants (I), 5
228 million inh. (S)). However, both Italy and Scotland were divided into two main regions (see Fig. 2).
229 This was made to avoid excessive fragmentation of the database into several small regional sub-sets
230 not ensuring statistical representation of the area, given also the length of the sample of segments
231 inquired. The considered regions are defined as follows:

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

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

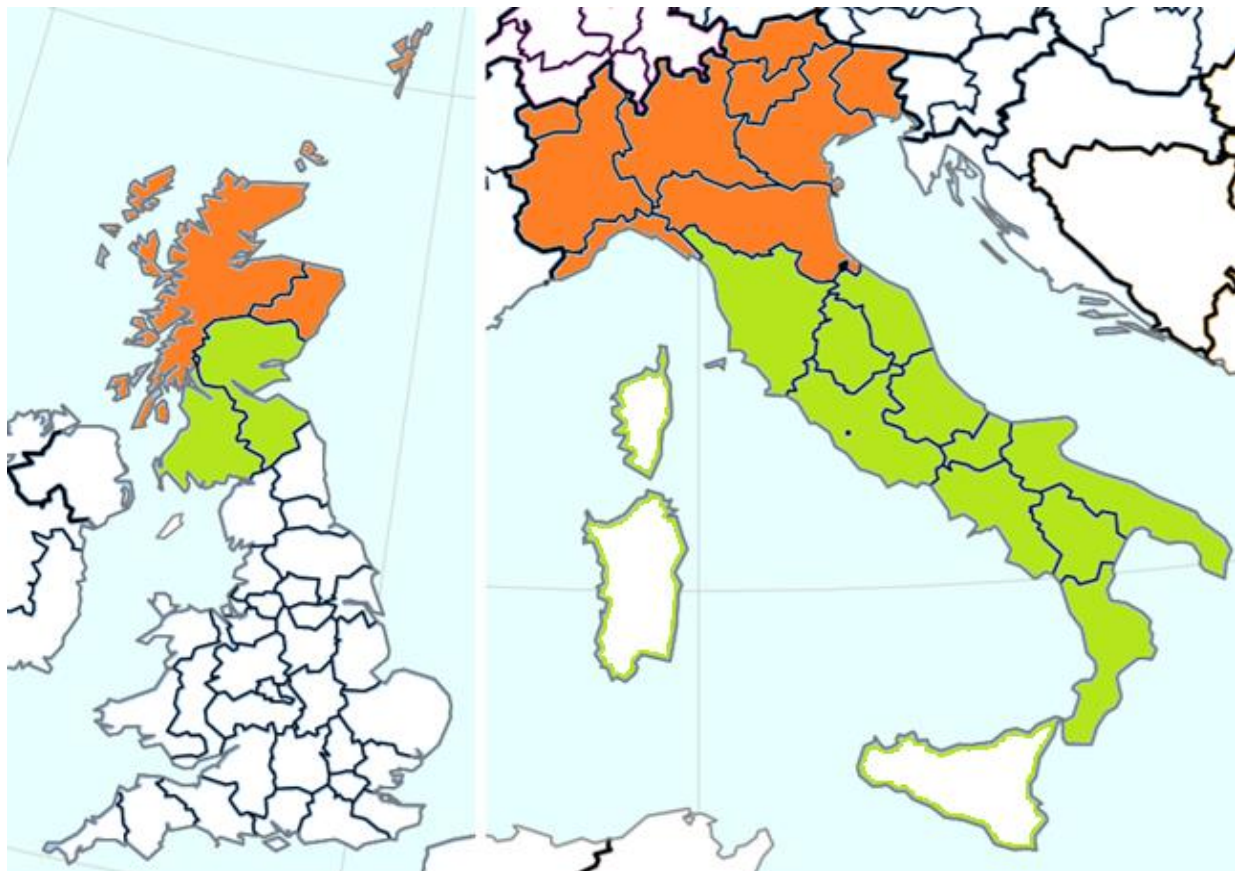
240 The two Scottish macro-regions were chosen based on the division into Lowlands and Highlands,
241 with historical and socio-cultural roots (e.g. Devine, 1979, Davidson, 2000). The macroscopic EU
242 NUTS 2 level classification (European Parliament and Council, 2003) divides Scotland into 4 regions:
243 East, South-West, North-East, Highlands/Islands. However, Scotland (far less wide than Italy) was
244 divided into two regions as well as Italy. Hence, Eastern and South-Western NUTS regions were
245 grouped into a “Lowlands” macro-region. Since North-Eastern Scotland is small and less densely
246 populated than the other Southern areas, it was grouped with the adjoining Highlands and Islands
247 NUTS region into a “Highlands” macro-region. As can be noted from Table 2, the division of
248 Scotland into Highlands (North) and Lowlands (South), based on traditional historic classifications,

249 is justified by geographic (i.e. population and population density) and infrastructural differences
 250 (variable “density of motorways” in Table 2), rather than other socio-economic comparisons.

251 *Table 2. Geographic and socio-economic variables for Italy and Scotland (data source:*
 252 <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

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



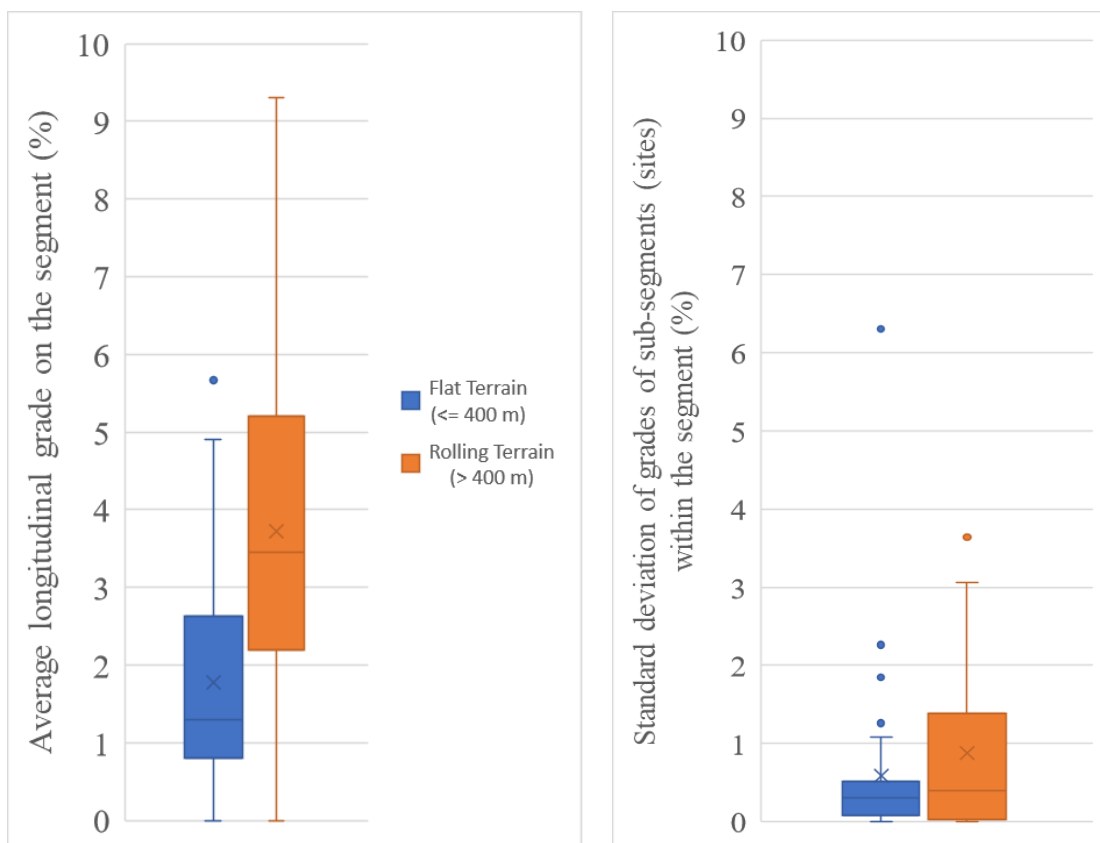
255
 256 *Figure 2. Map of regions considered in this study. Left: Scotland (“Highlands” in orange;*
 257 *“Lowland” in green). Right: Italy (Northern Italy in orange; Central/Southern Italy in green). Based*
 258 *on: <http://ec.europa.eu/eurostat/web/nuts/nuts-maps-.pdf>.*

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

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

265 For the Italian dataset, road sites were classified into: flat and rolling terrain (the latter is the most
266 widespread in Italy) (Colonna et al., 2016a). In the cited study, a binary terrain class (flat or rolling)
267 was assigned to each road site according to the average terrain elevation above/below the site.
268 Mountainous terrains were not present in the database. The elevation threshold between flat and
269 rolling terrains was set to 400 m above mean sea level. This value was previously identified as an
270 indicative limit beyond which the alignments of the secondary roads inquired are highly influenced
271 by surrounding terrains, through exploration of the road segments in the sample (Colonna et al.,
272 2016a). In this regard, the differences between the average gradients of segments and their variation
273 within the segment are shown in Figure 3. Boxplots clearly show how the two populations of gradients
274 above and below the 400 m selected threshold are different. Vertical alignments are more varying
275 and gradients are significantly steeper in the “rolling” than in the “flat” terrain class.

276



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

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

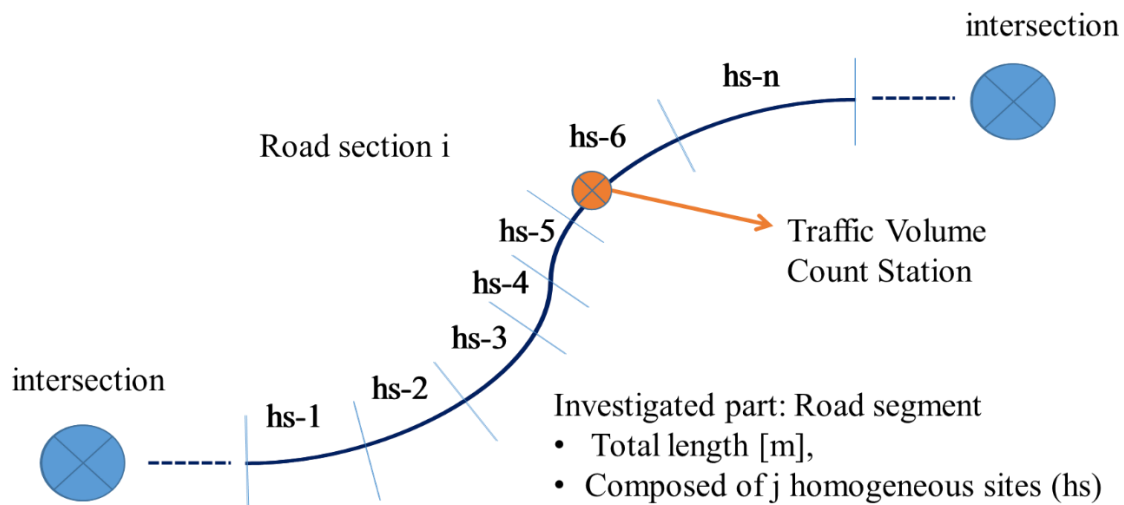
282 **2.3.2 Other variables**

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

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

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

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



312
 313 *Figure 4. Graphical scheme of road section and homogeneous sites.*

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

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

323

324 **2.4 Calibration procedure**

325 The performed calibration of a transferred SPF (TF) adopts: 1) the HSM (AASHTO, 2010) model for
326 two-lane rural roads as base reference SPF; 2) the calibration procedure described in the HSM for
327 transferring SPFs to different jurisdictions, (considering also improvements proposed by Lord et al.,
328 2016); 3) a procedure aimed at assessing the reliability of calibration (Bahar and Hauer, 2014).

329

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

- 333 • 30-50 homogeneous road sites;
- 334 • 100 accidents/year over the total sample of sites;
- 335 • 3 recent years of accident data.

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

341

342 For what concerns the minimum number of accidents, these are total accidents. Since fatal and injury
343 accident data are often more reliable than total accident data (or the only available), a sample
344 composed of a number slightly minor than 100 fatal+injury accidents per year may be sufficient (e.g.
345 Sacchi et al., 2012). The Italian database is composed of 422 fatal injury accidents over the period
346 2008-2012 (84.4 fatal+injury accidents/year). Hence, the requirement is deemed to be met for the
347 Italian case, and not for the Scottish case (101 fatal+injury accidents in the period: 2012-2014, 33.7
348 fatal+injury accidents/year). However, based on the information included in the accident database
349 investigated and their descriptions, the fatal+injury Italian and Scottish were equated to, namely,
350 KAB accidents (Colonna et al., 2016a; Cafiso et al., 2012) and KABC accidents (which account
351 namely for about 18 % and 32 % of total accidents, according to HSM estimates). The reference scale
352 taken into account is the KABCO scale (K = Killed, A = Incapacitating injury, B = Non incapacitating
353 injury, C = Possible Injury, O = Property Damage Only, PDO), provided in the HSM (AASHTO,
354 2010). This means that the Scottish 101 fatal+injury accidents may correspond to 316 total accidents,
355 which could meet the HSM recommendations. However, given this uncertainty, which broadly affects
356 the significance of results obtained for specific subsets (regions and traffic ranges), the reliability
357 assessment of calibration results is fundamental.

358

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

371
 372 For the Italian dataset, 10,000 vehicles/day was identified as a threshold dividing traffic ranges
 373 (Colonna et al., 2016a). In fact, previous studies (Sacchi et al., 2012; La Torre et al., 2014) highlighted
 374 that the HSM SPF tends to underestimate crash frequencies for high-crash sites, roughly for AADT
 375 > 10,000. Whereas, the Scottish dataset is mainly composed of low-volume roads (mean AADT:
 376 approx. 2,000 vehi./day, and standard deviation comparable to the mean). Hence, due to the high
 377 differences in traffic volumes of the two samples, the same Italian threshold was not deemed usable.
 378 Hence, it was set to 2,000 vehi./day; as this is close to the mean value of the sample of segments. In
 379 this way, the variability of calibration factors with traffic was investigated for Scotland as well, in the
 380 range of the traffic volumes in the sample.

381
 382 The calibration output is a calibration factor C_x , obtained by dividing the total observed accidents (in
 383 this case fatal+injury accidents) on the considered segments by the predicted accidents on the same
 384 segments (through the application of the base HSM SPF, the appropriate percentage of accident
 385 severities, and the applicable CMFs to each segment, according to the collected variables):

$$387 \quad C_x = \frac{\sum_{all\ segments} Observed\ number\ of\ accidents}{\sum_{all\ segments} Predicted\ number\ of\ accidents\ (uncalibrated\ SPF)} \quad (1)$$

388
 389 The calibration procedure was applied for both Italy and Scotland, and for the different subsets
 390 considered (two regions and traffic ranges for each country). Hence, an overall factor and other
 391 specific calibration factors are obtained.

392
 393 The C_x factors obtained were assessed by using the approach proposed by Bahar and Hauer (2014).
 394 The reliability assessment is based on $cv\{C_x\}$ values (Bahar and Hauer, 2014), computed as follows.
 395 They represent an estimate of the coefficient of variation of the associated C_x factors:

$$397 \quad cv\{C_x\} = \frac{\sqrt{\sum_{j=1}^n N_{a,j} + k_j N_{a,j}^2}}{C_x (\sum_{j=1}^n N_{u,j})} \quad (2)$$

398 Where:

400 $N_{u,j}$ = uncalibrated predicted number of crashes for the segment j ;

401 $N_{a,j} = C_x N_{u,j}$ = calibrated predicted crashes for the segment j (replaceable by observed crashes);

402 k_j = over-dispersion parameter (indicating a variance greater than the mean) of the base HSM SPF.

403 Values of $cv\{C_x\}$ less than 0.20 may be related to accurate C_x estimates (Bahar and Hauer, 2014).
 404 Hence, this value is deemed as a good threshold for assessing the reliability of calibration factors.
 405

406
 407 The improved guidelines for HSM calibration studies (Lord et al., 2016) were also taken into account,
 408 which provide the minimum number of road sites for obtaining a given level of accuracy. This number
 409 depends on the coefficient of variation of the observed accidents in the sample. If this minimum
 410 number is not achieved at a sufficient confidence level, the LF estimation is advised. Moreover, the
 411 need for region-specific calibration factors is suggested as well when the following disequation is
 412 satisfied. Otherwise, the State-wide calibration factor may be deemed as sufficient.

413 ,

$$414 e_r = \left| \frac{\frac{N_{obs,R}}{L_{average,R} * N_{SPF,HSM}(AADT_{average,R})}}{\frac{N_{obs,S}}{L_{average,S} * N_{SPF,HSM}(AADT_{average,S})}} - 1 \right| > 0.10 \quad (3)$$

415 Where:

416 $N_{obs,R/S}$ = observed accidents in the Regional (R)/State-wide (S) sample of road sites;

417 $N_{SPF,HSM}(AADT_{average,R/S})$ = accidents predicted from the baseline HSM SPF as a function of the
 418 AADT over the Regional (R)/State-wide (S) sample of road sites;

419 $L_{average,R/S}$ = average segment length in the Regional (R)/State-wide (S) sample of road sites (km).
 420

421 Alternative recent approaches may have been used for the HSM calibration (see e.g. Srinivasan et al.,
 422 2016; Farid et al., 2018a,b). However, a simple calibration approach was preferred (AASHTO, 2010),
 423 to better stress the different predictive capabilities, if any, of two extreme alternatives: LF estimation
 424 and TF calibration. However, guidance from Bahar and Hauer (2014) and Lord et al. (2016) were
 425 taken into account, as previously indicated, to assess the results from the HSM calibration. Additional
 426 references for these selected criteria can be found in Geedipally et al. (2017); Shirazi et al. (2016a,b).
 427

428 2.5 Modelling techniques

429 Accident modelling is often conducted by applying General Linear Modelling (GLM) approaches
 430 (Lord and Mannering, 2010), more flexible than linear modelling. Accident counts resulted over-
 431 dispersed (variance greater than the mean), thus the GLM regression was conducted by assuming a
 432 Negative Binomial (NB) distribution of the errors, and a natural logarithmic link function (Hilbe,
 433 2011; Chatterjee and Simonoff, 2013). This approach is commonly used for developing LFs (see Lord
 434 and Mannering, 2010 for a list of studies) and specifically for two-way two-lane rural roads (e.g.
 435 Zhang and Ivan, 2005; Cafiso et al., 2010; Russo et al., 2016). Zero-inflated models could be also
 436 used in these cases, since accident counts are often widely populated of zeros. However, their
 437 application was criticized for highway safety purposes (see Lord et al., 2005b) and the percentages
 438 of zeros in the sample of yearly accident frequencies are about 50 % (Italy) and 60 % (Scotland).

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

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

$$444 E(Y) = \exp(\beta_0) * L^{\beta_1} * AADT^{\beta_2} * \exp\left(\sum_{i=3}^n \beta_i X_i\right) \quad (4)$$

445 Where:

446 $E(Y)$ = predicted number of (fatal+injury) accidents per year (accidents/year);

447 L = length of the segment (m);

448 $\beta_0, \beta_2, \dots, \beta_n$ = estimated coefficients of the regression (β_1 is set to 1);

449 X_3, X_4, \dots, X_n = regression variables considered, other than segment length and AADT: road width,
450 shoulder type, radius of curvature, curve ratio, slope, driveway density, RHR, region, elevation.

451

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

465

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

$$469 \quad MC = \left(\frac{1}{MR}\right)_{curved\ part} * CR + \left(\frac{1}{MR}\right)_{straight\ part} * (1 - CR) = \left(\frac{1}{MR}\right)_{curved\ part} * CR \quad (5)$$

470 Where:

471 MC = weighted mean of the segment curvature (1/km), equal to zero for straight segments;

472 MR = mean radius of curvature of the curved part of the road segment (km), set to infinity in straight
473 parts of segments, thus leading to eliminate the second term of the weighted mean.

474

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

476

477 *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	<i>Italy</i> : 0 – Paved, 1 – Mixed-Composite/Gravel, 2 – Turf <i>Scotland</i> : 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

478

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

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

490

491

3. RESULTS

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

3.1 Italian case study

493

494

495

3.1.1 Italian HSM Calibration

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

498

499

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 Italy	Overall	112*	1.66	0.15	Yes (0.23)
	< 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	-

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

504 All calibration coefficients in Table 4 are reliable (Bahar and Hauer, 2014), except for low traffic in
505 Northern Italy and for high traffic in Southern/Central Italy. For some coefficients, including the
506 overall factor, the estimated equivalent “confidence levels” (Lord et al., 2016) are around 70 %, based
507 on the number of segments in the sample. This may justify HSM calibration instead of SPF

508 development. In particular, regional coefficients are advised for both the macro-regions considered
 509 (e_r values > threshold indicated in Eq. 3), especially for Northern Italy ($e_r = 0.23$).

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

517 A regional effect can be noted in the outputs of HSM calibration. The overall factor for Northern Italy
 518 ($C_x = 1.66$) is considerably higher than for Centre-South Italy ($C_x = 1.29$), and indeed a regional
 519 calibration factor was deemed necessary based on Lord et al. (2016). However, this difference may
 520 be attributed to the high percentage of high traffic sites for Northern Italy, considerably higher than
 521 the respective sites for Centre-South Italy. The higher traffic volumes for Northern Italian sites may
 522 have led to the notably high C_x for Northern Italy. Hence, pairwise comparisons between regions
 523 should be made by differentiating for traffic ranges. When comparing low traffic ranges, a notable
 524 difference emerges between Northern ($C_x = 1.39$) and Centre-South Italy ($C_x = 1.16$). However, the
 525 reliability of the Northern Italian low-volume coefficient is deemed questionable ($cv\{C_x\} = 0.22$).
 526 Whereas, when comparing high traffic ranges, no consistent differences may be noted (North: $C_x =$
 527 1.73; Centre-South: $C_x = 1.81$).

528 3.1.2 Local Safety Performance Function: Italy

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

532 *Table 5 – NB model parameters and goodness of fit measures for the Italian SPF, with p-values and*
 533 *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	ϑ
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)

534

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

543 Whereas, the following variables did not result statistically significant in the model development at
 544 the chosen significance level ($p = 0.10$): total road width, longitudinal slopes, driveway density, RHR,

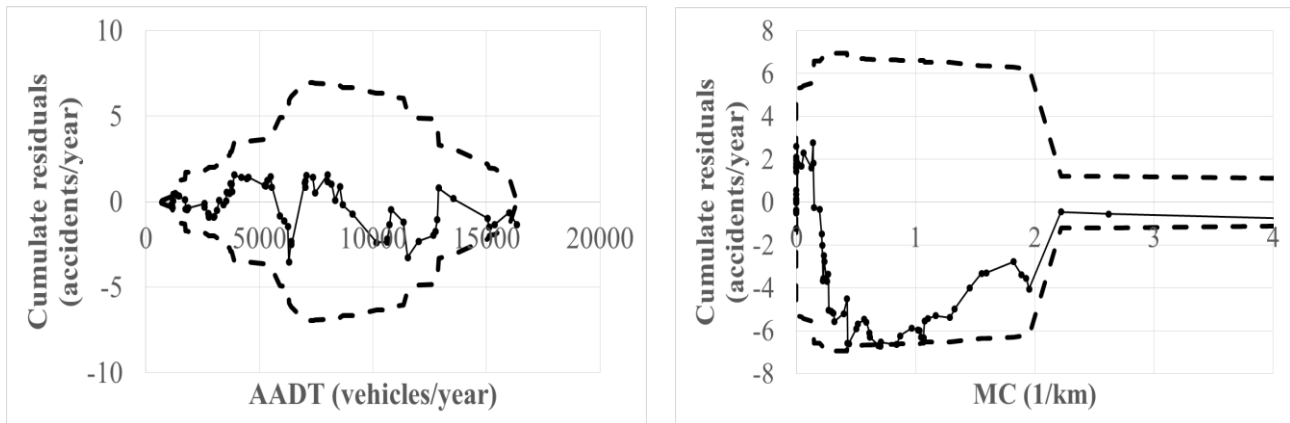
545 elevation. The variable region resulted statistically significant at the defined confidence level in the
 546 alternative model IT(A) reported in Table 6 indeed (as well as the longitudinal slope i , positively
 547 related to the accident frequency). However, the AIC value associated to the model IT(A) is greater
 548 than the corresponding value in Table 5 and thus, for this reason, the latter model was selected.
 549 However, given the research questions of this article, it is important to note that region may be
 550 considered as a significant variable in an alternative accident prediction model. Taking into account
 551 the model IT(A) in Table 6, Central-Southern Italy is associated to less accidents than Northern Italy,
 552 other variables being equal.

553

554 *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

555



556

557

558 *Figure 5. CURE plots for the Italian model (IT) related to the variables AADT and MC. Dashed lines*
 559 *represent the positive and negative two standard deviations ($\pm 2\sigma$).*

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

569

570 3.2 Scottish case study

571

572 3.2.1 Scottish HSM Calibration

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

575

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

Variable: Region	AADT Ranges	Number of Sites	Cx	cv[Cx]	Need for regional Cx (er) (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	-

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

580
 581 Most calibration coefficients presented in Table 7 may be deemed reliable (Bahar and Hauer, 2014),
 582 except for the Highlands factors differentiated for traffic ranges and the Lowlands factor for traffic
 583 volumes < 2,000 (subsets having the smallest number of sites). The sample of sites considered (even
 584 if comparable with the Italian ones) lead to estimated “confidence levels” of calibration < 70 %, due
 585 to less observed accidents, for which SPF development would be preferable (Lord et al., 2016). A
 586 regional coefficient would not be needed for both the two regions considered.

587 From the analysis of data in Table 7, the HSM SPF generally overestimates accident frequencies for
 588 Scottish two-lane rural roads (the overall and most of the other Cx factors are < 1). There is a notable
 589 difference between traffic ranges: Cx are considerably higher for low traffic volumes (< 2,000) than
 590 high volumes. This result is valid for the overall estimate (i.e. Cx = 1.20 for low volume sites and Cx
 591 = 0.48 for high volume sites) and even disaggregating data regionally. Hence, the overestimation
 592 effect of the HSM SPF (Cx < 1) is amplified for traffic volumes > 2,000 (associated to low Cx values).
 593 The most reliable Cx factors showing low cv{Cx} values are those obtained for the overall estimate
 594 and the first-level classification in regions and traffic ranges (i.e. not combining regions with traffic
 595 ranges). The Scottish calibration does not highlight any significant regional effect. The overall factor
 596 for the Lowlands (Cx = 0.75) is comparable to the Highlands (Cx = 0.66), as expected from the
 597 assessment procedure (no need for determining regional factors, based on Lord et al., 2016). This
 598 similarity can be noted even disaggregating according to the different traffic ranges.

599 **3.2.2 Local Safety Performance Function: Scotland**

600 The statistical parameters related to the fitted Scottish SPF are presented in Table 8, together with the
 601 over-dispersion parameters ϑ . The model satisfactorily fits accident data, according to goodness-of-
 602 fit measures. However, the pseudo-R² is considerably lower than the Italian model, and the over-
 603 dispersion parameter is greater.

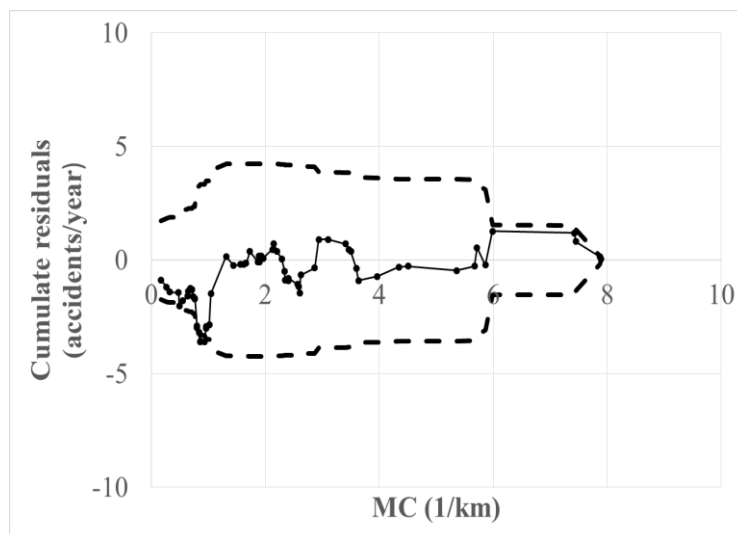
604
 605 The variables included in the model are: shoulder type and weighted curvature. They are all
 606 significant at the chosen significance level ($p = 0.10$). Shoulders made of turf (ST = 1) are negatively
 607 related to the accident frequency (i.e. less accidents in presence of turf shoulders), with respect to
 608 paved/mixed shoulders (reference condition: ST = 0). Weighted curvature is positively related to the
 609 accident frequency, such as in the Italian case.

610 The following variables did not result statistically significant at the chosen significance level ($p =$
 611 0.10): AADT, longitudinal slope, total road width, driveway density, RHR, region. Moreover, the
 612 analysis of the CURE plot in Fig. 6 reveals that the chosen model functional form is appropriate for
 613 what concerns the MC variable, with cumulate residuals oscillating around zero.

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

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

616
617



618
619

620 *Figure 6. CURE plot for the Scottish model (SC) related to the variable MC. Dashed lines represent*
 621 *the positive and negative two standard deviations ($\pm 2\sigma$).*

622
623

4. DISCUSSION

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

4.1 Calibration studies versus local SPF development

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

629 Calibration studies may be less demanding than LF estimations (especially if calibrations are
 630 conducted on base models including only some variables, e.g. traffic volumes, differently than the
 631 HSM calibration procedure, requiring several variables) and they may be conducted by non-experts
 632 through specific operational guidelines. However, the number of possible variables to take into
 633 account while conducting calibrations is some way limited by the necessary sample size for each
 634 combination of the considered variables. In fact, the reliability of a calibration factor may increase
 635 with the sample size, and minimum number of sites are suggested for calibration procedures
 636 (AASHTO, 2010; Lord et al., 2016). In this case, traffic, regions and the combinations of traffic

637 ranges and regions were considered as detailed disaggregation of the TF calibration study (i.e. a
638 calibration factor was derived for each combination of these variables). This means that several other
639 categories may have been considered, by further disaggregating the sample in small samples (e.g.
640 variables considered in the LF estimation: road width, curves, etc.).

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

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

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

670 For what concerns the general specific predictive capabilities of the calibrated TFs and estimated LFs
671 in this study, they are assessed based on computed residuals (difference between observed and
672 predicted values of yearly accident frequencies). To reveal possible significant improvements in the
673 prediction, residuals were computed for each of the subsets considered for calibration (overall,
674 regionally divided, classified into traffic ranges, classified into regions and traffic ranges). To allow
675 the comparison between different calibrated TFs and estimated LFs, the synthetic measure: MAD
676 (Mean Absolute Deviation) was used (such as in previous studies, see Oh et al., 2003; Sacchi et al.,
677 2012; La Torre et al., 2014). It is obtained as the sum of the absolute residuals computed for each
678 segment in the sample, divided by the number of segments. The closer the MAD index is to zero, the
679 more the prediction is accurate. The obtained MADs are reported in the following Table 9.

680 *Table 9 – Comparison of the Mean Absolute Deviation (MAD) [accidents/year] for the calibrated*
 681 *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

682
 683 An improvement in the prediction is noted for LFs with respect to calibrated HSM SPFs for both the
 684 Italian and Scottish case studies. An improvement is also noted if different regional and traffic subsets
 685 leading to specific calibration factors are considered, with respect to an overall calibration factor. As
 686 expected, the most relevant prediction improvement is noted while comparing MAD indexes of the
 687 locally developed SPF with the calibrated SPF. Paired t-tests were carried out to check the
 688 significance of the difference of the average MAD of corresponding calibrated and local SPFs. At the
 689 5 % significance level no statistically significant difference was detected.

690 This further result has several implications in light of the aims of this study. In fact, it is important to
 691 note that even if the prediction capabilities of estimated LFs are greater than those of calibrated TFs
 692 (overall and disaggregated), the differences are not statistically significant. This means that the effort
 693 of developing a novel SPF, based on the same sample which can be used for HSM calibration, may
 694 be not justified by a significant prediction improvement. Even if this conclusion is solely based on
 695 the two case studies considered and the associated samples of road segments, it may have important
 696 practical consequence. In this sense, it should be also noted that the LFs developed in this study are
 697 based on small sample sizes and small sample means of observed accidents. This may lead to biased
 698 estimations, including unreliable over-dispersion parameters, which may severely influence the
 699 expected accidents resulting from the application of the Empirical Bayesian (EB) method (Lord,
 700 2006). Hence, the development of local SPFs may be justified only in case of very large sample size,
 701 far greater than those required for HSM calibration, and in presence of several road and traffic
 702 variables collected. All these circumstances may lead to reliable and robust SPFs, which may
 703 significantly improve prediction capabilities with respect to simple calibrations. Otherwise, a detailed
 704 TF calibration (i.e. by at least considering the variability of traffic ranges) may represent a possible
 705 trade-off between computational, time and cost efforts and the reliability of results.

706 **4.2 Influence of geographic variability on crash predictions**

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

712 The regional variability does not add significant explanations of the accident frequency as a result of
 713 the Scottish LF. In fact, region was not a significant variable included in the final model. Whereas,
 714 in the Italian study, while terrain was not a significant predictor, the region variable was included in
 715 a model alternative to the model associated with the lowest AIC measure. If the model in Table 6
 716 would be used for accident prediction, estimates for Central-Southern Italy should be multiplied by
 717 $\exp(\beta_{REG})$, that is about 20 % smaller than predictions for Northern Italy, other conditions being equal.
 718 However, the final Italian model selected does not include the regional variable, but rather curvature
 719 and shoulder types, due to the associated improvement in the AIC score. The selection of model in

720 Table 5 is not only due to merely computational considerations. In fact, while regional classifications
721 may not be strongly influential on accident predictions, the influence of curvature is widely
722 documented (see e.g. Abdel-Aty and Radwan, 2000; Elvik, 2013b). Thus, the model in Table 5
723 (including curvature but excluding regions) was definitely preferred.

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

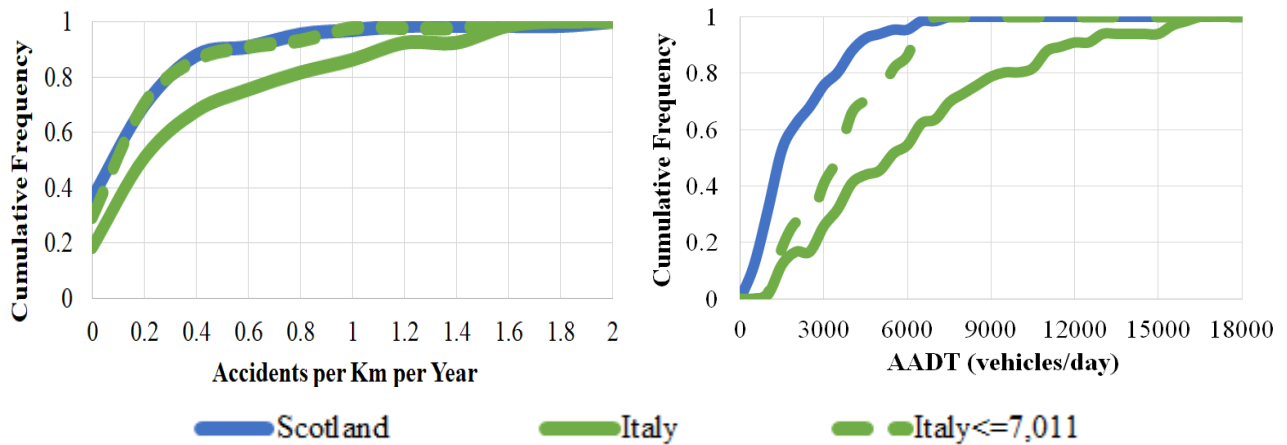
739 **4.3 Geographic variability of accident predictors: Italy versus Scotland**

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

744 A remarkable difference between the two case studies is the role of traffic volume in explaining
745 accidents. Traffic volume is often the most influential variable in predicting accident frequency. This
746 is confirmed in the Italian study, but not in the Scottish one. The exclusion of the traffic variable from
747 the Scottish SPF may seem surprising. However, the mean AADT for the Scottish sites is 2,048
748 vehi./day (st. dev.: 1,620 vehi./day); while the mean AADT for the Italian sites is 6,506 vehi./day (st.
749 dev.: 4,269 vehi./day), thus having a wider spectrum of traffic volumes. The difference in the road
750 networks of the two territories has contributed to the high discrepancy in traffic volumes. Secondary
751 Italian roads have mean traffic volumes greater than secondary Scottish roads, considering also that
752 Scottish two-lane “A” class rural roads (likely with more traffic than secondary roads) were excluded
753 from the database, because they belong to the primary network. However, it could be interesting to
754 compare the accidents-traffic relationship for the same traffic volume interval ($\leq 7,011$, maximum
755 Scottish volume). The cumulative frequencies of both accidents and traffic volumes are reported in
756 Figure 7, considering both database, and the Italian database with comparable volumes ($\leq 7,011$).

757 As can be noted in Fig. 7, both Scotland and Italy (for the same low-volume traffic range: $\leq 7,011$
758 vehi./day) exhibit a relevant frequency of zero-count sites (30-40 %), and a similar distribution of the
759 cumulative frequency of accidents/km (Fig. 7). However, when it comes to traffic volumes, there is
760 a notable difference between Italy and Scotland. Scottish volumes are heavily skewed on a very-low
761 volume (i.e. approx.. 40 % of sites have AADT $\leq 1,500$, and 75 % of sites have AADT $\leq 3,000$),
762 while Italian sites are not. This may have affected the search for a satisfactorily fitting accidents-

763 traffic curve. To note, an attempt Italian model fitted by considering only sites having $AADT \leq 7,011$
 764 still revealed traffic volume as a significant variable, even if with β_{AADT} close to 1, instead of > 1 .
 765



766
 767 *Figure 7 – Cumulative fatal+injury accident frequencies and traffic volumes for Italy and Scotland*
 768

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

- 771 • In case of a sample of secondary two-lane road sites having low traffic volumes and also
 772 skewed to very-low volumes, the accident frequency may not significantly be dependent on
 773 the amount of traffic volumes (as found for Scotland). This could be explained by the very
 774 low number of interactions between vehicles in the traffic flow, and most of the accidents may
 775 be single-vehicle accidents (e.g. run-off-road). This should be confirmed by future studies
 776 conducted on sites with AADT similar to the Scottish sites. Moreover, in this case, as
 777 explained above, different calibration factors obtained for different traffic ranges (as in this
 778 case, using 2,000 as a threshold) may be unreliable even if statistically valid (Table 4).
- 779 • In case of a sample of two-lane road sites having a wide spectrum of traffic volumes as the
 780 Italian ones, the relationship between accident and traffic was found to be more than linear
 781 ($\beta_{AADT} \sim 1.4$). However, when separating only sites with $AADT \leq 7,011$ (comparable to the
 782 Scottish ones), the inferred accidents-traffic relationship becomes approximately linear.
 783 Hence, a linear relationship as the one in the HSM (AASHTO, 2010) may only be valid for
 784 low traffic volumes (approximately $< 10,000$, see Sacchi et al., 2012). Hence, in case of sites
 785 with widely varying traffic volumes, different traffic ranges should be considered if only
 786 calibration is conducted. In this way, the nature of the non-linear accidents-traffic relationship
 787 may be captured also in a calibration procedure (see Table 4, $C_{X, \geq 10,000} \gg C_{X, < 10,000}$).
 788

789 The effect of curvature is strongly related to accident frequencies, as found in previous studies (e.g.
 790 Abdel-Aty and Radwan, 2000; Elvik, 2013b). This is valid for both Italy and Scotland. The effect of
 791 curvature is more evident on Italian than on Scottish sites, by comparing the β_{MC} coefficients. This
 792 may be explained by the nature of road sites considered. Italian sites have mean CR: 0.14 (st. dev.:
 793 0.15), mean radius of curvature: 295 m (st. dev.: 195 m), while Scottish sites have mean CR: 0.55 (st.
 794 dev.: 0.26), mean radius of curvature: 349 m (st. dev.: 275 m). Hence, mean radii of curvature of
 795 curved segments are similar, while the percentages of curved sites on the segment (CR) are not.
 796 Scottish segments are notably more winding than the Italian ones. The small segment curvature may
 797 lead Scottish drivers to select lower speeds and this, in turn, may result in lower accident risks (Aarts
 798 and Van Schagen, 2006; Elvik, 2013a). The reduced accident risk may also be due to the smaller

799 skidding risk at low speeds (Colonna et al., 2016b). On the other hand, Italian drivers may select
800 higher speeds on the sample of road sites due to the low percentage of curves. Because of the higher
801 Scottish segment curve ratio, the mean speed differential between consecutive segments and curves
802 (especially if sharp) for Scottish drivers may likely be lower than the corresponding Italian drivers'
803 speed differential. The inclusion of variables which attempt at capturing operating speeds and speed
804 differences (see e.g. Cafiso et al., 2010) may have helped in revealing those differences related to
805 samples of roads with different importance. Since it was not possible to derive those variables from
806 the dataset inquired, further research on the regional variability of accident predictions should
807 consider also speed variables. Local operating speed models (see e.g. Discetti et al., 2011) may help
808 for this aim, even if relying on a predicted operating speed as a base variable for SPFs may lead to an
809 increase in both the uncertainty and the unreliability of results.

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

825

826

5. CONCLUSIONS

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

- 833 • A trans-national variability of accident predictions was noted between Italy and Scotland. This
834 was largely associated to the different nature of the two two-lane road networks. The
835 representative Scottish road sites present lower traffic volumes and design features (i.e. more
836 curves, unpaved shoulders, etc.) than the Italian sample of sites. This affected the modelling
837 stage, revealing a not significant influence of traffic on Scottish accidents. The highlighted
838 result and the possible existence of traffic volume thresholds below which the influence of
839 traffic decreases should be verified in future studies for very-low volume roads.
- 840 • An inner variability of accident predictions was not found in the Scottish case, while it was
841 individuated in the Italian case study (in both calibration and the intermediate stages of SPF
842 development and selection). However, as explained in the text, a weak regional variability
843 may rather hide the influence of other variables. Anyway, the finally selected Italian model

844 did not include region as a significant predictor. This may lead to conclude that time and costs
845 necessary for considering geographic variability of crash predictions among administrative
846 boundaries may be saved, by prioritizing other variables. The homogeneity of road design
847 standards among countries may be prevalent on local differences (e.g. drivers' behaviour).
848 This was evident in Scotland, while further studies could be needed in the Italian case.

- 849 • Calibration procedures (especially those accounting only for some variables) may be
850 inexpensive and easier than LF estimation. However, even statistically significant calibration
851 factors may be “false positives” when checked against results of a comprehensive SPF, such
852 as the differences between traffic ranges and regions in this study. On the other hand, LF
853 estimations based on the same sample size required for TF calibrations may only slightly
854 improve the predictive capabilities of a simple TF calibration, as revealed in this study. Hence,
855 when sufficiently large and statistically representative sample size, and the related detailed
856 datasets of road/traffic features are not available, the efforts for estimating a new LF could be
857 saved and the TF calibration could be a good compromise (e.g. for practitioners, when LFs
858 are not available).
- 859 • The segment curvature and the shoulder types were revealed as significant crash predictors in
860 both the Italian and Scottish models, even with some local differences, attributed to the
861 different importance of roads and their possible influence on speeds (which were not modelled
862 in this study). Road width, elevation, roadside hazard, driveway density and longitudinal
863 slopes resulted not statistically significant accident predictors in both models.

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

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

882 **ACKNOWLEDGEMENTS**

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

886

887 **REFERENCES**

- 888 Aarts, L., & Van Schagen, I. (2006). Driving speed and the risk of road crashes: A review. *Accident*
889 *Analysis & Prevention*, 38(2), 215-224.
- 890 Abdel-Aty, M. A., & Radwan, A. E. (2000). Modeling traffic accident occurrence and
891 involvement. *Accident Analysis & Prevention*, 32(5), 633-642.
- 892 American Association of State Highway and Transportation Officials (AASHTO) (2010). *Highway*
893 *Safety Manual*. Washington, DC.
- 894 Bahar, G. B., & Hauer, E. NCHRP 20-07 (332) (2014) *Users' Guide to Develop Highway Safety*
895 *Manual Safety Performance Function Calibration Factors*, Transportation Research Board,
896 Washington, D.C.
- 897 Bauer, K. M., & Harwood, D. W. (2000). *Statistical Models of At-Grade Intersection Accidents*.
898 Addendum (Report No. FHWA-RD-99-094). United States Federal Highway Administration.
- 899 Brimley, B., Saito, M., & Schultz, G. (2012). Calibration of Highway Safety Manual safety
900 performance function: development of new models for rural two-lane two-way highways.
901 *Transportation Research Record: Journal of the Transportation Research Board*, (2279), 82-89.
- 902 Cafiso, S., Di Graziano, A., Di Silvestro, G., La Cava, G., & Persaud, B. (2010). Development of
903 comprehensive accident models for two-lane rural highways using exposure, geometry, consistency
904 and context variables. *Accident Analysis & Prevention*, 42(4), 1072-1079.
- 905 Cafiso, S., Di Silvestro, G., & Di Guardo, G. (2012) Application of Highway Safety Manual to Italian
906 divided multilane highways. *Procedia-Social and Behavioral Sciences*, 53, 911-920.
- 907 Chatterjee, S., & Simonoff, J. S. (2013). *Handbook of regression analysis*. John Wiley & Sons.
- 908 Colonna, P., Berloco, N., Intini, P., Perruccio, A., Ranieri, V., & Vitucci, V. (2016a). Variability of
909 the Calibration Factors of The HSM Safety Performance Functions with Traffic, Region and Terrain.
910 The case of the Italian rural two-lane undivided road network. *Compendium of Papers of the 95th*
911 *Annual Meeting of the Transportation Research Board* (No. 16-3413).
- 912 Colonna, P., Berloco, N., Intini, P., Perruccio, A., & Ranieri, V. (2016b). Evaluating skidding risk of
913 a road layout for all types of vehicles. *Transportation Research Record: Journal of the Transportation*
914 *Research Board*, (2591), 94-102.
- 915 Colonna, P., Intini, P., Berloco, N., & Ranieri, V. (2018). Integrated American-European protocol for
916 safety interventions on existing two-lane rural roads. *European transport research review*, 10(1), 5.
- 917 Davidson, N. (2000). *The origins of Scottish nationhood*. Pluto Press.
- 918 Devine, T. M. (1979). Temporary migration and the Scottish Highlands in the nineteenth century. *The*
919 *Economic History Review*, 32(3), 344-359.
- 920 Discetti, P., Dell'Acqua, G., & Lamberti, R. (2011). Models of operating speeds for low-volume
921 roads. *Transportation Research Record: Journal of the Transportation Research Board*, 2203(1),
922 219-225.
- 923 Donnell, E., Gayah, V. & Jovanis, P. (2014). *Safety Performance Functions*. Pennsylvania
924 Department of Transportation. Report no. FHWA-PA-2014-007-PSU WO 1.

- 925 Elvik, R. (2013a). A re-parameterisation of the Power Model of the relationship between the speed
926 of traffic and the number of accidents and accident victims. *Accident Analysis & Prevention*, 50, 854-
927 860.
- 928 Elvik, R. (2013b). International transferability of accident modification functions for horizontal
929 curves. *Accident Analysis & Prevention*, 59, 487-496.
- 930 European Parliament and Council (2003). *Regulation (EC) No 1059/2003 of the European*
931 *Parliament and of the Council of 26 May 2003 on the establishment of a common classification of*
932 *territorial units for statistics (NUTS)*.
- 933 European Union. Eurostat. <http://ec.europa.eu/eurostat/web/nuts/nuts-maps-.pdf>. Last accessed on
934 June, 6th, 2018. <http://ec.europa.eu/eurostat/data/database>. Last accessed on August, 8th, 2018.
- 935 European Union (2015). Regions in the European Union. Nomenclature of territorial units for
936 statistics. NUTS 2013/EU-28. Eurostat. Manuals and Guidelines.
- 937 Farid, A., Abdel-Aty, M., Lee, J., Eluru, N., & Wang, J. H. (2016). Exploring the transferability of
938 safety performance functions. *Accident Analysis & Prevention*, 94, 143-152.
- 939 Farid, A., Abdel-Aty, M., & Lee, J. (2018a). A new approach for calibrating safety performance
940 functions. *Accident Analysis & Prevention*, 119, 188-194.
- 941 Farid, A., Abdel-Aty, M., & Lee, J. (2018b). Transferring and calibrating safety performance
942 functions among multiple states. *Accident Analysis & Prevention*, 117, 276-287.
- 943 Farid, A., Abdel-Aty, M., & Lee, J. (2019). Comparative analysis of multiple techniques for
944 developing and transferring safety performance functions. *Accident Analysis & Prevention*, 122, 85-
945 98.
- 946 Garach, L., de Oña, J., López, G., & Baena, L. (2016). Development of safety performance functions
947 for Spanish two-lane rural highways on flat terrain. *Accident Analysis & Prevention*, 95, 250-265.
- 948 Garber, N. J., Haas, P. R., & Gosse, C. (2010). *Development of safety performance functions for two-*
949 *lane roads maintained by the Virginia Department of Transportation* (No. FHWA/VTRC 10-R25).
- 950 Geedipally, S.R., Shirazi, M., Lord, D. (2017). *Exploring the need for region-specific calibration*
951 *factors*. Transportation Research Record, 2636, 73–79.
- 952 Gomes, S. V., Geedipally, S. R., & Lord, D. (2012). Estimating the safety performance of urban
953 intersections in Lisbon, Portugal. *Safety science*, 50(9), 1732-1739.
- 954 Gomes, M. M., Pirdavani, A., Brijs, T., & Pitombo, C. S. (2019). Assessing the impacts of enriched
955 information on crash prediction performance. *Accident Analysis & Prevention*, 122, 162-171.
- 956 González, M. P. S., Sotos, F. E., & Ponce, Á. T. (2018). Impact of provincial characteristics on the
957 number of traffic accident victims on interurban roads in Spain. *Accident Analysis & Prevention*.
- 958 Gooch, J. P., Gayah, V. V., & Donnell, E. T. (2018). Safety performance functions for horizontal
959 curves and tangents on two lane, two way rural roads. *Accident Analysis & Prevention*, 120, 28-37.
- 960 Greibe, P. (2003). Accident prediction models for urban roads. *Accident Analysis &*
961 *Prevention*, 35(2), 273-285.

- 962 Hauer, E. (1999). *Safety in geometric design standards*. University of Toronto, Department of Civil
963 Engineering.
- 964 Hauer, E., & Bamfo, J. (1997). Two tools for finding what function links the dependent variable to
965 the explanatory variables. In *Proceedings of the ICTCT 1997 Conference*, Lund, Sweden.
- 966 Hauer, E., Bonneson, J., Council, F., Srinivasan, R., Zegeer, C. (2012). Crash modification factors:
967 foundational issues. *Transportation Research Record: Journal of the Transportation Research*
968 *Board*, 2279, 67–74.
- 969 Hauer, E., Persaud, B. (1997). *Safety analysis of roadway geometric and ancillary features*. Research
970 Report. Transportation Association of Canada.
- 971 Hauer, E. (2015). *The art of regression modeling in road safety* (Vol. 38). New York: Springer.
- 972 Hilbe, J. M. (2011). *Negative binomial regression*. Cambridge University Press.
- 973 ISTAT. Istituto Nazionale di Statistica. [Italian National Institute of Statistics]. <http://dati.istat.it/>.
974 Last accessed on June, 05th, 2018.
- 975 Italian Ministry of Infrastructures and Transport (2001). Functional and Geometric standards for road
976 design. D.M. 6792.
- 977 Jackman, S. (2017). Political Science Computational Laboratory, ‘pscl’ package v.1.5.2.
- 978 Kononov, J., & Allery, B. (2003). Level of service of safety: Conceptual blueprint and analytical
979 framework. *Transportation Research Record: Journal of the Transportation Research Board*, 1840,
980 57-66.
- 981 La Torre, F., Domenichini, L., Corsi, F., & Fanfani, F. (2014). Transferability of the Highway Safety
982 Manual Freeway Model to the Italian Motorway Network. *Transportation Research Record: Journal*
983 *of the Transportation Research Board*, 2435, 61-71.
- 984 Lee, J., Abdel-Aty, M., & Cai, Q. (2017). Intersection crash prediction modeling with macro-level
985 data from various geographic units. *Accident Analysis & Prevention*, 102, 213-226.
- 986 Li, L., Gayah, V. V., & Donnell, E. T. (2017). Development of regionalized SPFs for two-lane rural
987 roads in Pennsylvania. *Accident Analysis & Prevention*, 108, 343-353.
- 988 Liu, J., Khattak, A. J., & Wali, B. (2017). Do safety performance functions used for predicting crash
989 frequency vary across space? Applying geographically weighted regressions to account for spatial
990 heterogeneity. *Accident Analysis & Prevention*, 109, 132-142.
- 991 Lord, D., & Mannering, F. (2010). The statistical analysis of crash-frequency data: a review and
992 assessment of methodological alternatives. *Transportation Research Part A: Policy and*
993 *Practice*, 44(5), 291-305.
- 994 Lord, D., Manar, A., & Vizioli, A. (2005a). Modeling crash-flow-density and crash-flow-V/C ratio
995 relationships for rural and urban freeway segments. *Accident Analysis & Prevention*, 37(1), 185-199.
- 996 Lord, D., Washington, S. P., & Ivan, J. N. (2005b). Poisson, Poisson-gamma and zero-inflated
997 regression models of motor vehicle crashes: balancing statistical fit and theory. *Accident Analysis &*
998 *Prevention*, 37(1), 35-46.

- 999 Lord, D. (2006). Modeling motor vehicle crashes using Poisson-gamma models: Examining the
1000 effects of low sample mean values and small sample size on the estimation of the fixed dispersion
1001 parameter. *Accident Analysis & Prevention*, 38 (4), 751-766.
- 1002 Lord, D., Geedipally, S. R., & Shirazi, M. (2016). *Improved guidelines for estimating the Highway
1003 safety manual calibration factors* (No. ATLAS-2015-10). University Transportation Centers
1004 Program (US).
- 1005 Martens, M. H., Compte, S., & Kaptein, N. A. (1997). *The effects of road design on speed behaviour:
1006 a literature review*. Deliverable D1 (Report 2.3.1), MASTER.
- 1007 McCullagh, P. (1984). Generalized linear models. *European Journal of Operational Research*, 16(3),
1008 285-292.
- 1009 Myers, R. H., Montgomery, D. C., Vining, G. G., & Robinson, T. J. (2012). *Generalized linear
1010 models: with applications in engineering and the sciences*. John Wiley & Sons.
- 1011 Oh, J., Lyon, C., Washington, S., Persaud, B., & Bared, J. (2003). Validation of FHWA crash models
1012 for rural intersections: Lessons learned. *Transportation Research 26 Record: Journal of the
1013 Transportation Research Board*, (1840), pp. 41-49.
- 1014 Oh, J., Washington, S.P., Nam, D. (2006). Accident prediction model for railway highway interfaces.
1015 *Accident Analysis and Prevention* 38 (2), 346–356.
- 1016 Persaud, B., Lord, D., & Palmisano, J. (2002). Calibration and transferability of accident prediction
1017 models for urban intersections. *Transportation Research Record: Journal of the Transportation
1018 Research Board*, (1784), 57-64.
- 1019 PRACT Project. Predicting Road ACCidents - a Transferable methodology across Europe.
1020 <http://www.practproject.eu/>. Last accessed on June, 05th, 2018.
- 1021 Qin, X., Ivan, J. N., Ravishanker, N., & Liu, J. (2005). Hierarchical Bayesian Estimation of Safety
1022 Performance Functions for Two-lane Highways using Markov Chain Monte Carlo modeling. *Journal
1023 of Transportation Engineering*, 131(5): 345-351.
- 1024 Russo, F., Busiello, M., & Dell, G. (2016). Safety performance functions for crash severity on
1025 undivided rural roads. *Accident Analysis & Prevention*, 93, 75-91.
- 1026 Sacchi, E., Persaud, B., & Bassani, M. (2012). Assessing international transferability of highway
1027 safety manual crash prediction algorithm and its components. *Transportation Research Record:
1028 Journal of the Transportation Research Board*, 2279, 90-98.
- 1029 Sawalha, Z., & Sayed, T. (2006). Transferability of accident prediction models. *Safety science*, 44(3),
1030 209-219.
- 1031 Scottish Office for National Statistics. Detailed information on the administrative structure within
1032 Scotland. [https://www.ons.gov.uk/methodology/geography/ukgeographies/administrativegeography/
1033 scotland](https://www.ons.gov.uk/methodology/geography/ukgeographies/administrativegeography/scotland). Last accessed on June, 05th, 2018.
- 1034 Shin, H. S., Dadvar, S., & Lee, Y. J. (2015) Results and Lessons from the Local Calibration Process
1035 of the Highway Safety Manual for the State of Maryland. In *Transportation Research Board 94th
1036 Annual Meeting* (No. 15-4643).

- 1037 Shirazi, M., Geedipally, S. R., Lord, D. (2016)a. A procedure to determine when safety performance
1038 functions should be recalibrated. *Journal of Transportation Safety & Security*, 9(4), 459-467.
- 1039 Shirazi, M., Lord, D., Geedipally, S. R. (2016)b. Sample-size guidelines for recalibrating crash
1040 prediction models: Recommendations for the Highway Safety Manual. *Accident Analysis &
1041 Prevention*, 93, 160-168.
- 1042 Srinivasan, R., & Carter, D. L. (2011). *Development of safety performance functions for North
1043 Carolina*. Chapel Hill, NC: North Carolina Department of Transportation, Research and Analysis
1044 Group. (Report No. FHWA/NC/2010-09).
- 1045 Srinivasan, R., Colety, M., Bahar, G., Crowther, B., & Farmen, M. (2016). Estimation of calibration
1046 functions for predicting crashes on rural two-lane roads in Arizona. *Transportation Research Record:
1047 Journal of the Transportation Research Board*, (2583), 17-24.
- 1048 Sun, X., Li, Y., Magri, D., & Shirazi, H. (2006). Application of highway safety manual draft chapter:
1049 Louisiana experience. *Transportation Research Record: Journal of the Transportation Research
1050 Board*, 1950, 55-64.
- 1051 Tarko, A. P. (2006). Calibration of safety prediction models for planning transportation networks.
1052 *Transportation Research Record: Journal of the Transportation Research Board*, 1950(1), 83-91.
- 1053 Transport Scotland (2016). Scottish Transport Statistics. A National Statistics Publication for
1054 Scotland. 2016 Edition. No. 35.
- 1055 UK Department for Transport (2012). Guidance on road classification and the primary route network.
- 1056 Venables, W. N. & Ripley, B. D. (2002) *Modern Applied Statistics with S*. Fourth Edition. Springer,
1057 New York. ISBN 0-387-95457-0.
- 1058 Washington, S. P., Karlaftis, M. G., & Mannering, F. (2010). *Statistical and econometric methods for
1059 transportation data analysis*. Chapman and Hall/CRC.
- 1060 Xie, F., Gladhill, K., Dixon, K., & Monsere, C. (2011). Calibration of Highway Safety Manual
1061 Predictive Models for Oregon State Highways. *Transportation Research Record: Journal of the
1062 Transportation Research Board*, 2241, 19-28.
- 1063 Yannis, G., Dragomanovits, A., Laiou, A., Richter, T., Ruhl, S., La Torre, F, Domenichini, L.,
1064 Graham, D., Karathodorou, N. & Li, H. (2016). Use of accident prediction models in road safety
1065 management - an international inquiry. *Transportation Research Procedia*, 14, 4257-4266.
- 1066 Zegeer, C. V., & Deacon, J. A. (1987). *Effect of lane width, shoulder width, and shoulder type on
1067 highway safety*. State of the art report, 6, 1-21.
- 1068 Zhang, C., & Ivan, J. (2005). Effects of geometric characteristics on head-on crash incidence on two-
1069 lane roads in Connecticut. *Transportation Research Record: Journal of the Transportation Research
1070 Board*, 1908, 159-164.