Reliability Engineering and System Safety 188 (2019) 125–132

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

journal homepage: www.elsevier.com/locate/ress

Investigating safety performance of the SAFESTAR system for route-based curve treatment

Salvatore Cafiso[a], Carmelo D'Agostino[b,*], Mariusz Kiec[c]

[a] Department of Civil Engineering & Architecture, University of Catania, Via Santa Sofia 64, I-95125 Catania, Italy
[b] Transport & Roads Division, Department of Technology and Society, Faculty of Engineering, LTH, Lund University, P.O. Box 118, 221 00 Lund, Sweden
[c] Faculty of Civil Engineering, Cracow University of Technology, Warszawska 24, 31-155 Cracow, Poland

ARTICLE INFO

Keywords:
Road safety
Horizontal alignment
Road sign
SAFESTAR system
Crash Modification Factor
Safety Performance Function

ABSTRACT

In order to improve curve safety, warning signs are used to alert drivers to unexpected changes in horizontal alignment and speed variations.

While all countries have standards and guidelines for signs, markings and delineations to ensure a uniform system, they do not use them in the same way and choices and combinations may differ. SAFESTAR project promoted a method for establishing uniform curve signs on two lane rural roads to force drivers to adjust their speed to the actual risk. Although widely applied around the world and producing speed reduction benefits, as evidenced in the literature, the reliability in terms of statistical safety significance of the method has not been well established.

The goal of this paper is to assess the safety fundamentals of SAFESTAR's risk ranking system and safety performance of the proposed signing schemes. Crash data and road characteristics from two Polish regions are used to carry out sound statistical analysis. Building on data from treated and untreated sites, the analysis first looks at relative changes in Crash Rates for different risk categories. Next, Crash Modification Factors are estimated. Results are interesting and suggest that system needs revision, both in terms of risk classification and on how treatments are implemented.

1. Introduction

Curves play a fundamental role in the safety performance of road alignment [1]. For that reason, road administrations pay particular attention to design, delineation and signing criteria. Because a major road safety improvement by retrofitting on the existing road network is not possible, given the extent and costs of the works, other solutions may be considered. In order to improve curve safety and reduce crashes and crash severity, warning signs are used to alert drivers to changes in the geometry that may not be apparent or expected [2–4]. In many countries, various systems of horizontal curve signing and marking are used [5–9] with the same aim to improve traffic safety and reduce speed. The European SAFESTAR project [10] proposed a method for selecting the most appropriate combination of marking, signing and delineation based on risk category (from A to E) of curves in two lane rural roads, ensuring consistency of signing of curves in the network. In SAFESTAR, higher risk categories require more treatments (guide posts, chevron markers and profiled edge road markings). While many different countries and their road authorities use approaches similar to those proposed in SAFESTAR, the safety benefits of those treatments in terms of crash reduction are not well understood because they have not been assessed with a complete and sound methodology offering statistical significance [11].

The effectiveness of the SAFESTAR system was mainly assessed on the basis of speed behavior [12] rather than on the actual safety performance.

In the original work, a full-scale test on 13 curves (6 Danish and 7 French) with risk category B, C and D showed limited speed reductions of 2 km/h or more in approach speed at 8 sites and of 2 km/h or more in curve speed at 7 sites out of 13 [13]. The authors stressed that the model could only be used outside Denmark and France, if calibrated for the prevailing national conditions.

The Portuguese Road Administration adopted the SAFESTAR approach in 2005. An FH inconsistency factor was defined. It weights the expected increase in accident risk using a measure of variation in kinetic energy required to drive along a curve. Curves are divided into five consistency classes, depending on their FH, the values of speed reduction on approaches and the expected deceleration rate [14]. Trial

* Corresponding author.
E-mail addresses: dcafiso@dica.unict.it (S. Cafiso), carmelo.dagostino@tft.lth.se (C. D'Agostino), mkiec@pk.edu.pl (M. Kiec).

https://doi.org/10.1016/j.ress.2019.03.028
Received 29 June 2018; Received in revised form 4 March 2019; Accepted 10 March 2019
Available online 11 March 2019
0951-8320/

tests on a 170 km stretch of a Portuguese main highway showed that the application of the signing system for curves has improved some aspects of driver behavior, including lower speeds (−5 to −11 km/h, in the 85th percentile of speed on curves) [14]. The SAFESTAR model is currently applied in the UK by TRL [15].

Guidelines for selecting curve-related traffic control devices were proposed by Bonneson et al. [8] in 2007, based largely on the existing practices of many transportation agencies in the US. Application of the guidelines begins with a determination of the curve's severity category based on a variation of the 85th percentile speed between tangent and curve. In this case the introduction of V_{85} for speeds either for tangents or curves as well as the relationship between risk category and speed differential, is based on kinematic assumption similar to SAFESTAR [8].

The region of Malopolska in Poland introduced SAFESTAR in 2002 as a pilot study, measuring approach speed before and after the treatment on 8 curves (B and C risk category). The new marking and signing have changed some aspects of driver behavior, including lower speeds (−2.6 km/h) for C risk category in the 85th percentile of approach speed, and a slight increase in the 85th percentile of approach speed for B risk category (+0.6 km/h). The impact of new marking and signing on road safety was not statistically analyzed [17].

Despite the extended practice, the actual crash reduction as a result of SAFESTAR has not been scientifically analyzed in Poland or in other countries where similar procedures are currently applied [11]. Similarly, some differences exist in the setting of the upper and lower limits of the risk categories. Thresholds are based on values of variance in kinetic energy, maximum deceleration and speed reduction that are not always correlated to the effective changes in the expected crash rate. Moreover, the transferability of classification criteria among different countries is not trivial and should be verified, as well.

Taking this as reference, the goal of the paper is to assess the safety fundamentals of SAFESTAR's risk rating system and the consequent safety performance of the signing schemes proposed as a countermeasure. The case study involves crash data and road characteristics from two regions in Poland (treated and untreated sites) providing the basis for a statistical analysis of the system's safety performance.

2. Literature review

A literature review of curve signing schemes and their effect on safety shows a variation in crash reduction ranging from 5% to 40% as shown in Table 1 [11], not fully representing, anyhow, the effects of the SAFESTAR approach.

A study reported by Elvik et al. [1] on the effects of chevron markings have been found to reduce accidents by 32% (95% CI [−59; +13]). On British motorways, chevrons have been found to increase headways, and both collisions and single vehicle accidents have been reduced.

Elvik et al. [1] reported that combined measures on delineation seem to have larger effects on accidents than each of the measures separately as reported in Table 2, for which the combined effects were studied, not considering the product of each single Crash Modification Factor (CMF) which would have produced a larger accident reduction. This is particularly true if the countermeasures target the same crash type, such as combined signing on curves. Therefore, multiplying several CMFs of complementary treatments is likely to overestimate the combined effect [18].

Table 2
Effects of combined measures on crash reduction related to delineation treatments [1] (all the estimate reductions are related to all accident types and injury severity, except for "*" for which the severity was unspecified).

Combined measures	Percentage change in the number of accidents	
	Estimate reduction	95% confidence interval
Edge line and pavement markers*	−47	(−66; −18)
Edge lines and directional markings in curves	−19	(−46; +23)
Pavement markers and directional markings in curves	−45	(−58; −28)
Edge line and center line	−24	(−35; −11)
Edge line, center line and delineator posts	−45	(−56; −32)

Similar results are included in CMF clearinghouse website. Recent references indicate improvement of road safety after implementation of combined measures on delineation. Montella [6] estimated CMF for implementation of: a combination of chevron signs, curve warning signs, and sequential flashing beacons (0.524), chevron signs and curve warning signs (0.556) and chevron signs only (0.63) on horizontal curves. Srinivasan et al. [19] estimated, significantly lower CMFs for implementation of chevron signs on horizontal rural roads curves (0.84) and new fluorescent curve signs or upgrade existing curve signs to fluorescent sheeting (0.75).

Another approach to develop an engineering crash risk assessment model for ranking of curves was proposed by Jurewicz et al. [9] in Australia. Three risk categories were created: low, medium and high, calculated based on curve direction, change in approach speed, pavement width, road grade. For each curve risk category, a CMF was estimated by combining existing CMFs of different treatments.

For Czech National Roads, Ambros et al. [16] proposed a combined risk assessment of curves based on three steps: a) speed consistency, b) tangent length and following curve radius, c) radius of two consecutive curves. For classes from A to C, an incremental application of traffic control devices was proposed (solid centerline, warning signs, chevron, advisory speed, double solid centerline). For class D reconstruction is recommended. Expected safety performance of the different treatments are not reported.

Lamm [2] analyzed curves with 3 incremental equipment levels (marking and signs) concluding that level 2 or even level 3 improved traffic safety, however not to a level which would correspond to good design practices according to safety criteria I, II and III.

Despite the fact that the available CMFs refer to delineation treatments of curves, those values cannot be directly applied to the SAFESTAR system for several reasons. The criteria for selecting the delineation treatment are not consistent with SAFESTAR and, even if implementing several countermeasures could be more effective, it is unlikely that the full effect of each countermeasure would be obtained when implemented concurrently with others. Ultimately, literature review shows that there is a lack of studies about safety performance of the SAFESTAR or equivalent approaches.

Table 1
Summary of findings about crash reduction related to curve delineation treatments [9].

Treatment type	Brief description	Crash reduction	Speed reduction
Advance warning signs	Used in advance of curves to raise attention level and slow motorists.	25%	Unknown
Chevron alignment markers (CAMs)	Used to indicate presence and severity of curves.	30%	3.5 km/h
Speed advisory signs	Sometimes used to help indicate the comfortable travelling speed (and hence the severity) of a curve.	40%	Unknown
Other delineation devices	Includes guide posts, line marking, pavement markers, etc. to provide additional guidance for safe roadway negotiation.	5–20%	May increase

3. The safestar system

The literature review [7] indicates that there are three feasible measures of curve driving severity, based on a differential: speed, energy and friction. The energy differential approach was used in the SAFESTAR EU project developed in 2002 under the EU Transport Research program [8]. The method aims at "*establishing uniformity for the signing of curves on two-lane rural roads to force the drivers to adjust their speed to the degree of risk in the curve sufficiently enough before entering it*" [8].

In the energy approach, each curve is classified based on risk factors falling into 5 categories (A, B, C, D, E). Each category is defined for minimum and maximum values of the difference in the kinetic energy ΔE_{kin} between the speed in curve and the speed on the preceding tangent, defined by Eq. (1):

$$\Delta E_{kin} = c\ (V_{approach}^2 - V_{design}^2) \tag{1}$$

where

$V_{approach}$ is the 85th percentile of the operating speed (V85) on the approaching tangent in km/h;

V_{design} is the design speed in curve given by the well-known expression of Eq. (2) again in km/h:

$$V_{design} = \sqrt{127 \cdot R \cdot (f_r + e)} \tag{2}$$

where

c is the system constant

R is the radius of curve [m];

f_r is the side friction factor; and

e is the value of superelevation [%].

The model stems on the fact that it is the size of the "required reduction" of the kinetic energy that determines the risk of the curve, irrespective of the speed level. In other terms, at higher speeds a smaller speed reduction is required for the same level of risk. The risk category is thus based on the assumption that only small changes in the approach speed are acceptable at high speed levels, while larger changes are acceptable at low speed levels for curves of the same risk category. In the original study, the upper and lower limits of ΔE_{kin} for each category were calibrated for French and Danish conditions [13] (Fig. 1).

Category A represents curves where the speed differentials are low enough and drivers tend to reduce speed slightly. Category E, however, represents the sharpest curves where drivers will have to begin braking well before they reach the curve.

Once the risk of the curve has been identified, signs and markings for that curve are installed according to the risk category. The higher the risk category the more treatments are applied. These include advance curve warning signs, guide posts, chevron markers and ordinary or profiled road markings. Reconstruction of curve geometry is suggested for curves of risk category E, where the marking should only be used as a temporary measure.

SAFESTAR's warning sign system was gradually introduced in Poland from 2002 to 2010 as a regional low cost large-scale action to improve road safety. While the original goal of SAFESTAR risk classification remained unchanged, the signing system was adapted to Polish design standards. The main difference is that curve warning signs are also applied to category A and the installation of speed limits starts from category B (Fig. 2). In the original system, there are no curve warning signs in category A and a mandatory speed limit was set at 40 km/h only for categories D and E [9].

The approaching speed, depending on the road configuration (e.g. length of tangents, cross section width, gradient), must be defined based on actual speed or speed calculated from speed models separately for each direction. Because crash data are not specified per direction, the worst condition of risk class was assumed.

In 2010 the SAFESTAR system covered about 1000 km of the Małopolska region's roads. In other regions, the system is not mandatory.

The standard Polish guidelines for traffic signing require the use of the warning signs (Fig. 3) based on the radius and super elevation with guideposts on the curve depending on the radius. As regards chevrons, Polish guidelines are not uniform on that and suggest the use of chevrons on horizontal curves if the geometry is likely to be unexpected for drivers or if the radius is smaller than the road's standard radius and where deflection is high. All this makes curve marking and signing inconsistent and leaves drivers unable to understand and follow curve speed.

4. Data gathering and treatment

Because the main goal of the paper is to estimate the safety performance of the SAFESTAR approach applied in Poland, similar sites were selected from the region where the system was applied (treated

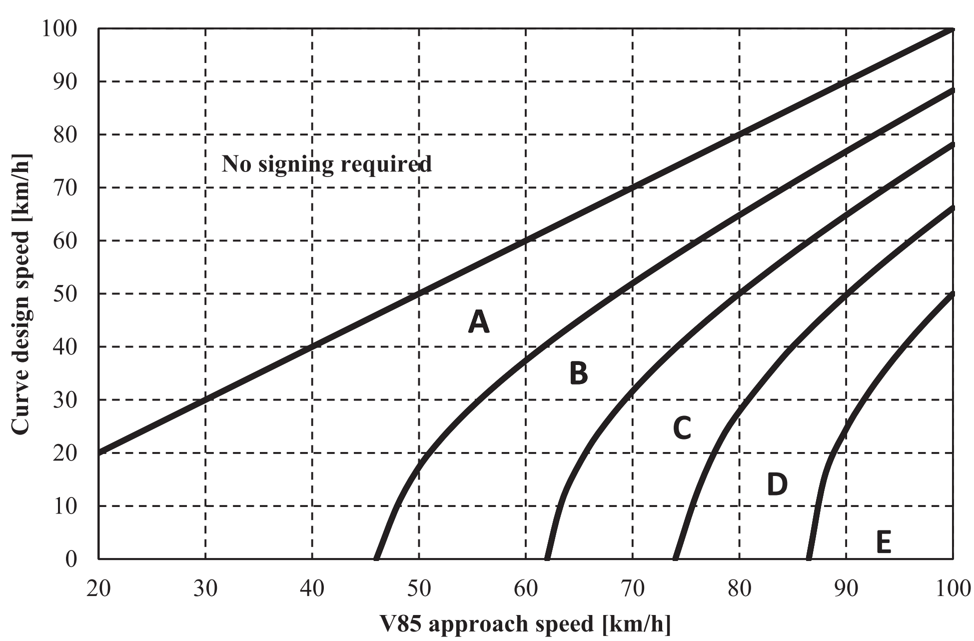

Fig. 1. SAFESTAR risk classification [8].

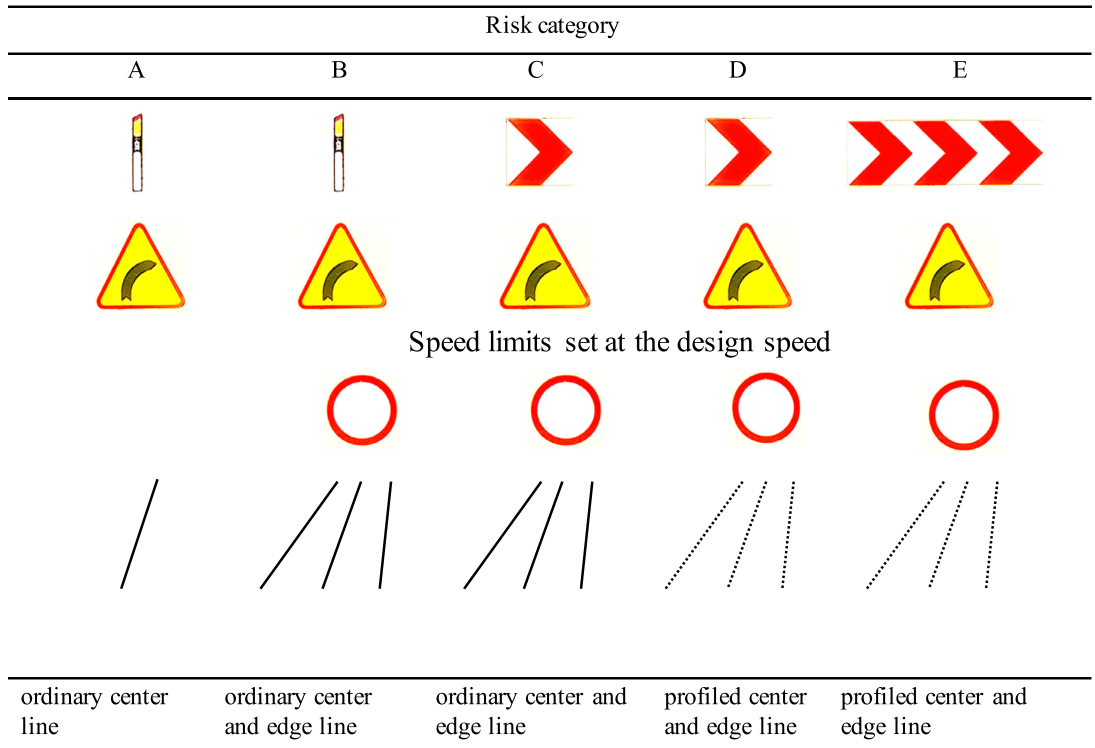

Fig. 2. Signing and marking for the different risk categories of the SAFESTAR system adopted in Poland.

sites) and from the region where it was not applied (untreated sites). The data set is composed of 523 km of treated roads and a comparison group of untreated roads of 227 km (Table 2).

Where treated roads are concerned, SAFESTAR's system (Fig. 2) was first introduced in 2002. The region with no treated sites followed the standard Polish guidelines for traffic signing (Fig. 3).

Each curve in the data set is part of two lane rural roads with similar section width (lanes of 3.0 ÷ 3.5 m and shoulder 0.75 ÷ 1.5 m), level and rolling terrain.

Table 3 shows the information available for each curve with a statistical summary of the data set composed by treated and untreated sites.

5. Calibration of models and estimation of CMFs

Since the SAFESTAR model was calibrated using French and Danish data and a direct correlation between risk category and crash occurrence was not originally performed, an analysis on the safety performance of circular curves with different risk categories was carried out on Polish data to check if the road sections falling into different risk categories actually show different crash rates.

This section presents the methodology and the results of the analysis regarding the assessment of safety performance of the SAFESTAR System, by way of a classification of crash rate variability among sites with different categories of risk, and an estimate of a Crash Modification Factor for the treatment.

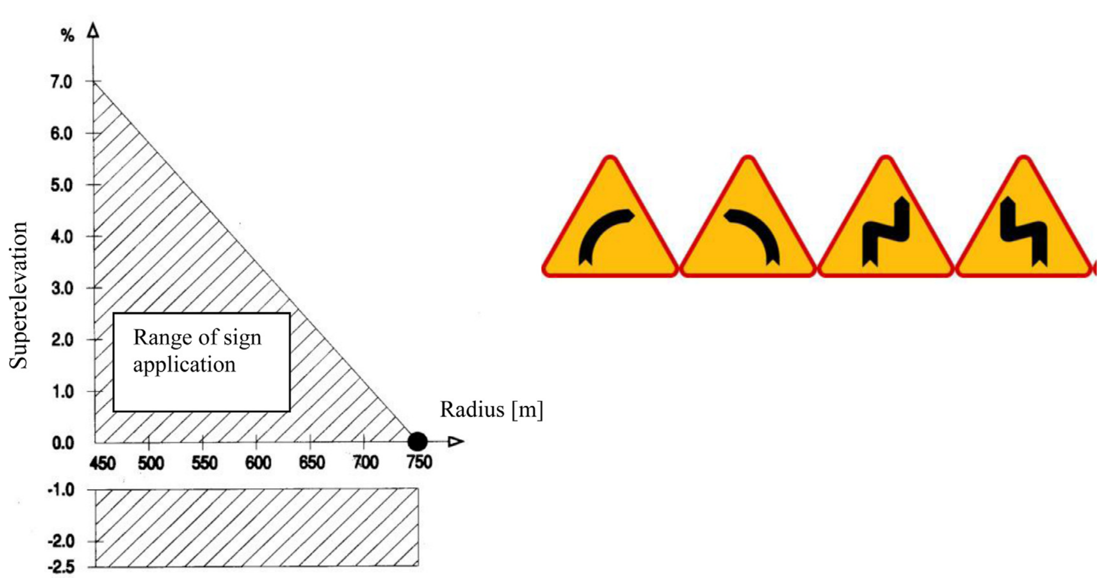

Fig. 3. Obligatory warning signs under the Polish guidelines.

Table 3
Summary statistics of variables used in the elaboration.

Variable	Sym.	Unit	Treated/Untreated	N	Min	Max	Av.	St. Dev
Category A	Cat	Categorical Variable	T	529	.	.	.	.
			U	294	.	.	.	.
Category B			T	520	.	.	.	.
			U	94	.	.	.	.
Category C			T	294	.	.	.	.
			U	82	.	.	.	.
Category D			T	40	.	.	.	.
			U	10	.	.	.	.
Category E			T	6	.	.	.	.
			U	0	.	.	.	.
Fatal + Injury Target Crashes (Σ5 years)*	E [y]	Dependent Variable	T	566	0	7	0.41	0.84
			U	138	0	4	0.32	0.67
Length	L	m	T	.	12.21	737.23	75.54	56.10
			U	.	13.51	565.98	103.40	58.85
Radius**	R	m	T	.	20.00	1950.00	219.27	187.40
			U	.	22.00	1440.00	256.30	166.87
Deflection angle	Def	degree	T	.	3.00	184.00	28.99	24.65
			U	.	5.00	155.00	29.16	19.56
Curvature Change Rate***	CCR	gon/m	T	.	32.67	3185.00	290.51	339.91
			U	.	44.24	2895.45	248.54	381.73
Av. Annual Daily Traffic	AADT	Vehicle per day	T	.	1659	15,112	5664.85	2907.08
			U	.	2345	13,374	5778.45	3227.41

* Crash data include only severe crashes (fatal + injury) and run-off-road, head on, rear-end, angle and sideswipe (same and opposite direction) typologies.
** The curve radius [m] represents the best fitting of alignment to a circular curve.
*** CCR is a measure of the curvature change rate of the alignment defined in [2].

5.1. Safety performance of curves

To estimate the expected crash rate among the different risk categories, an SPF was calibrated on the untreated sites, using as categorical variable the risk category of the SAFESTAR classification. Equation form 3 was used for that model:

$$E(Y) = \exp(\alpha) \times L \times AADT \times \exp(\delta_i \cdot Cat) \tag{3}$$

where:

E (Y) is the predicted average 5 year crash frequency (fatal + injury);

L is the length of road curves [km];

AADT is the average value 5 years' traffic flow [veh/day];

Cat is the categorical variable related to the SAFESTAR risk categories (A, B, C, D and E);

αregression term, δ_i are the regression terms of categorical variables *Cat*.

The base form of Eq. (3) was selected to better identify the average crash rate of road curves classified in SAFESTAR'S four categories of risk (A, B, C, D). Category E was not included due to a lack of data in the untreated sample. Length of curve *L* and AADT were included in model as offset variables (exponent = 1) to allow a direct comparison of the crash rate of different risk categories. Assuming that crashes are proportional to exposure is an approximation, but it is commonly adopted by HSM [20] (Length in this case). Table 4 shows the regression results.

To compare, a similar regression was calibrated on the treated sites (Table 4). In that model, curves in the D and E categories were combined due to the limited number of sections in both categories.

The regression coefficients have been used to estimate the "average" Crash Rate (*CR*) for each risk category of the treated and untreated sites by the following Eq. (4):

$$CR_i = exp(\alpha) * exp(\delta_i) \tag{4}$$

where:

α and δi are the estimate of the coefficients of Eq. (3) reported in Table 4.

CR is computed as crashes per million of vehicle-kilometers per day - Crash$*10^6$/ (365*AADT*km).

Results are shown in Fig. 4. It is clear that as the risk category increases, an increase in the average *CR* is expected in the untreated sites. The average crash rate on category D curves is almost seven times the average *CR* on category A curves.

Table 4
Value of regression parameters, standard error and *p*-value for the SPFs calibrated on the untreated and treated curves for each category defined by SAFESTAR system.

Parameter	Category	Estimate	Error	Chi-Square	Pr > ChiSq
Untreated sites					
Intercept	.	0.6146	0.7159	309.77	< 0.0001
Cat	A	− 1.9962	0.7253	7.58	0.003
Cat	B	− 1.5858	0.7484	4.49	|
Cat	C	− 1.18	0.7859	2.25	|
Cat	D	0	0	.	|
Dispersion	.	0.7699	0.3429	.	.
Treated sites					
Intercept	.	− 0.5015	0.3616	1567.4	< 0.0001
Cat	A	− 0.1672	0.3708	0.2	0.001
Cat	B	− 0.2922	0.3732	0.61	|
Cat	C	0.1389	0.3799	0.13	|
Cat	DE	0	0	.	|
Dispersion	.	0.9824	0.1629	.	.

Notes: variable were considered significant when *p*-value ≤ 0.05. The categorical variables were considered significant when both *p*-value of Type I and III analysis were ≤ 0.05.

As expected, that trend is strongly correlated to the increase in the kinetic energy (R^2 = 0.99) given by the difference between the approach and curve speeds as expressed by factor $\Delta E = (V^2_{approach} - V^2_{design})$. What was unexpected, is that the increase is not uniformly distributed among risk categories, because an increase of only 50% exists between categories A, B and C, with an abrupt step of 217% in CR of sites D when compared to sites of category C.

If only the treated sites are considered, the trend is flat and less statistically robust (R^2 = 0.40) showing no variation between categories A, B, C and D. When treated and untreated sites are compared, a clear difference in *CR* is observed only for sites of category D. To assess independence between the predicted crash rate in the treated and untreated sites, a *Z* test was performed on the values of mean and variance of the regression coefficients for each class. The results pointed out that

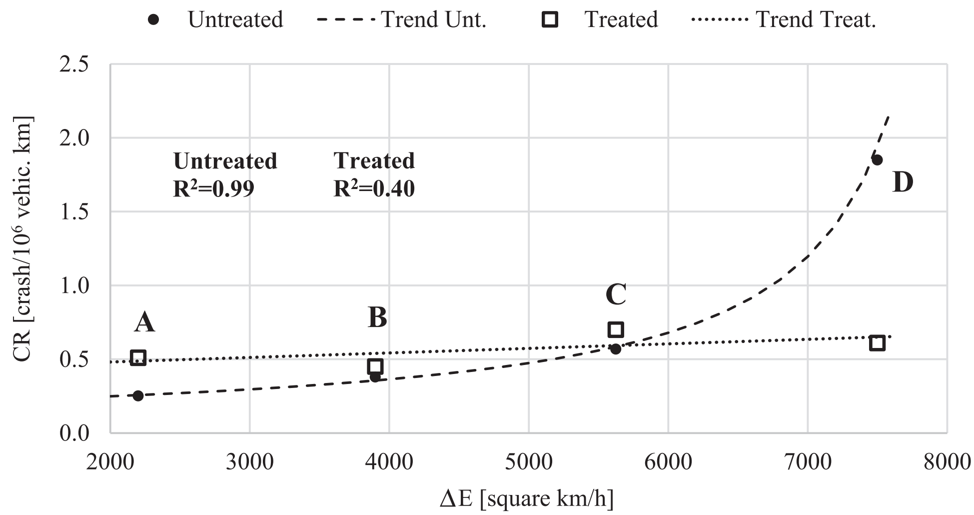

Fig. 4. Comparison of CR for different risk categories of treated and untreated curves.

only in the D class the treated sites resulted statically different than the untreated with a level of confidence of 95%.

Those results are not conclusive in terms of the CMF, because CRs were estimated separately on different datasets, with a simple model form and did not include other variables that would affect the estimation. However, these preliminary results should raise concerns in the application of the SAFESTAR system, either in terms of risk category identification or treatment effectiveness. The former would lead to improper site selection and the latter to ineffective crash reduction and budget allocation. In that framework, the evaluation of the expected crash reduction must be correctly estimated by calibrating the CMFs based on a comparison of treated and untreated sites, as reported further in the paper.

5.2. Crash modification factor (Cross sectional analysis)

A CMF is a multiplicative factor used to compute the change in the expected number of crashes after a given countermeasure is implemented. The expected number of crashes after the treatment is then computed by multiplying the CMF with the expected crash frequency without treatment.

A GLM regression model taking account of a Negative Binomial distribution of error, was calibrated using the entire dataset composed of treated and untreated sites.

The model form is reported in Eq. (5) and the results of calibration are contained in Table 5.

$$E(Y) = \exp(\alpha) \times L \times AADT^{\beta} \times \exp(\delta_i \cdot Cat) \times \exp(\gamma_1 \cdot Def + \gamma_2 \cdot CCR) \times \exp(\lambda \cdot Treated) \qquad (5)$$

where:

E (Y) is the predicted average 5-year crash frequency (fatal plus injury);

L is the length of road curves [m];

AADT is the average annual daily traffic [veh./day];

Cat is the categorical variable related to the SAFESTAR categories (A, B, C, D and E);

Treated is the categorical variable to identify treated or untreated sites (0, 1);

Def is the deflection angle of the Curve [deg], and *CCR* is the Curvature Change Rate [gon/m].

α, β and γ are regression terms of numerical variables;

δ_i and λ are the regression terms of categorical variables.

Variables for the model were selected using a stepwise methodology, i.e. by including only those variables whose *p*-value is higher than 0.05 and by minimizing the AIC value [27] (Akaike Information Criterion). The goodness of fitting for all the calibrated models was checked using cumulative residual (CURE) plots [28]. All the models show a suitable fit to the observed data with no drift outside the -2σ

Table 5
Value of regression parameters, standard error and *p*-value for the SPF calibrated on treated and untreated sites for the entire database and for each category defined by SAFESTAR system.

Parameter	Category	Estimate	Error	Chi-Square	Pr > ChiSq	CMF
Treated + Untreated sites for the entire dataset						
Intercept	.	−13.1801	0.8529	238.81	<0.0001	.
AADT	.	0.9896	0.0936	111.85	<0.0001	.
Cat	A	−0.7741	0.306	6.4	0.0034	.
Cat	B	−0.5261	0.3059	2.96	|	.
Cat	C	−0.0397	0.3103	0.02	|	.
Cat	DE	0	0	.	|	.
Def	.	−0.0055	0.0023	5.66	0.0174	.
CCR	.	0.0006	0.0001	22.97	<0.0001	.
Treated	1	−0.0953	0.1065	0.8	0.0123	0.91
Untreated	0	0	0	.	.	1.0
Dispersion	.	0.8266	0.1299			
A						
Intercept	.	−15.2059	1.2159	156.4	<0.0001	.
AADT	.	1.1495	0.137	70.41	<0.0001	.
Treated	1	0.1570	0.1366	1.92	0.009	1.17
Untreated	0	0	0	.	.	1.0
Def	.	−0.0118	0.0039	9.42	0.002	.
CCR	.	0.0007	0.0002	21.32	<0.0001	.
Dispersion	.	0.8075	0.1736	.	.	.
B						
Intercept	.	−16.5914	1.8128	83.77	<0.0001	.
AADT	.	0.9121	0.1574	33.57	<0.0001	.
Treated	1	−0.1795	0.2002	0.58	0.0463	0.84
Untreated	0	0	0	.	.	1.0
Def	.	−0.0079	0.0045	3.05	0.050	.
CCR	.	0.6148	0.1611	14.56	0.0001	.
Dispersion	.	0.5183	0.2123	.	.	.
C						
Intercept	.	−16.7138	2.1692	59.37	<0.0001	.
AADT	.	1.4016	0.2603	28.99	<0.0001	.
Treated	1	−0.16041	0.3372	0.17	0.050	0.85
Untreated	0	0	0	.	.	1.0
Radius	.	−0.0025	0.0017	2.11	0.014	.
Dispersion	.	0.959	0.3345	.	.	.

Notes: variable were considered significant when *p*-value ≤ 0.05. The categorical variables were considered significant when both *p*-value of Type I and III analysis were ≤ 0.05.
By definition CMF $= \exp(\lambda_{Treated})$.
Untreated is the comparison category with $\exp(\lambda_{Untreated}) = 1.0$.

boundaries.

The estimated CMF from the model coefficient is equal to 0.91 for all the treated sites which explains an average 9% reduction in crashes due to the SAFESTAR classification and curve treatment. Considering the standard error of the estimate, the level of confidence for which the upper limits of the CMF becomes equal to 1 is 85%, which practically means no changes in crash frequency after the implementation in all treated sites (categories from A to D).

The effectiveness of the treatment for curves in each risk category can be estimated by modeling the SPF on the treated and untreated sites belonging to one risk category only.

The model form is Eq. (6), with the same meaning of symbols as in Eq. (5), while regression results are reported in Table 5:

$$E(Y) = \exp(\alpha) \times L \times AADT^{\beta} \times \exp\left(\sum_{i=1}^{2} \gamma_i \cdot Var_i\right) \times \exp(\lambda \cdot Treated) \tag{6}$$

where:

*α, β*and*γi,* are regression terms (related to the multiplier variables); and

Treated the class variable in the regression related to the treated or untreated sites (the base condition), respectively.

Var_i are numerical variables included in the model that are the same for curves in categories A and B (*Def* and *CCR*) and Radius only [m] in C category.

D and E categories are not reported in the table because the SPFs calibrated on D and E categories (even if combined) did not have significant parameters due to the small sample size.

The results are interesting because they show that treatments on curves of category A are ineffective with a potential 17% increase in crash frequency. Treatments of curves in categories B and C show a similar average reduction in crash rate of about 15%. Again, when the standard error of the estimate is considered, the 95% upper limits of the CMFs becomes higher than 1.0, meaning no changes in crash frequency after implementation.

Overall results of CMF estimations are reported in Table 6.

In comparison with the preliminary results showed in the previous section of the paper (Fig. 4), categories B and C have not experienced an increase in crash occurrence but a 15% average reduction can be expected. Category A confirmed an unusual increase of 17% in the expected accident frequency.

6. Discussion on study design to develop CMFs

CMFs can be used by safety engineers in performance based design of highways to compare the safety effects of design alternatives, estimate the reduction in expected crash frequency before the implementation of various countermeasures in different locations and perform cost-benefit analysis in terms of crash effects [19,21–22].

Estimating the expected crashes without treatment is not a trivial task. It is not simply the number of observed crashes before or without treatment since this value could be higher or lower than expected due to regression-to-the-mean or to other confounding factors not directly addressed in the estimation. Other parameters that affect the safety of a facility, such as traffic volume, may change over time. Consequently, specific evaluation techniques are required to account for changes in order to estimate the "true" effects of safety improvements [18] avoiding naïve methodologies which are not able to take into account possible bias due to the random nature of crashes and the other confounding factors or time trend [23,24].

In that framework, the main practical concern in the CMF calibration is the collection and availability of enough data to ensure a proper statistical inference and a significant statistical result. As an alternative approach to crash data, surrogate measures of safety can also be used to evaluate the safety performance of a treatment [25,26]. Moreover, although the Empirical Bayes or the full Bayes before-after methodologies are considered the more appropriate study designs, different techniques can be applied to develop CMFs depending on data availability. Considering all the possible approaches in the estimation of the effects of treatment based on observed crashes, the methodology chosen for the analyses was targeted on the data typologies and availability [18,29]. More specifically, cross-sectional designs was selected because of insufficient number of instances where the countermeasure was applied in order to conduct a before-after study.

Specifically, CMFs derived from cross-sectional data are based on a single period, under the assumption that the ratio of average crash frequencies for sites with and without treatments is an estimate of the CMF. The main challenge in that approach is to collect data for a sufficient number of locations that are similar in all factors affecting crash risk. Hence, cross-sectional analyses are often accomplished through multiple variable regression models to account for other variables that may affect safety. The model is then used to estimate the CMF from the regression coefficient of the variable of interest. That value is a measure of crash change from a unit change in the specific variable or as an average variation in case of class statement.

Particularly data lacks of either sample size, or temporal period. In other terms there were not fully availability of curves and segments which can cover all the possible solutions explored by the SAFESTAR approach, but some of the category has not enough data to perform a robust regression analysis.

The sample size limitation arises in the estimation of the effects for each class of the SAFESTAR classification. Curves in the category E were not statistically significant due to the sample size, and they were aggregated with the category D, which from the SAFESTAR perspective have a different kinetic energy.

7. Conclusions

The SAFESTAR system was introduced in 2002 in the framework of an EU project and was initially applied in France and Denmark. The approach became very attractive for road administrations in different countries (e.g. UK, PL, Australia, US) because curves are marked, signed and delineated based on a standardized classification of risk category (from A to E), ensuring consistency of curve signing along the two-lane rural road network. Even if the risk classification and signing system are based on sound principles and consistency concepts, the actual safety effectiveness of the system and practical transferability in different countries was not analyzed in depth.

With data now available after the 2002 implementation of the SAFESTAR system in one of Poland's regions, a statistical analysis can be conducted to test the classification criteria and signing treatments against the actual crash history. The present research work dealt with some limitations in the data sample which, anyway, do not debar the study validity or the practical application can derive from it.

The results are of interest because they demonstrate that there is no gradual increase in crash rates among the 4 risk categories, with a slight difference of about 50% between the A, B, C categories and an abrupt increase (+217%) from 0.6 to 1.9 crashes per 10^6 vehicles/km from C to D category.

Moreover, the boundaries of category A are ambiguous because they include signing with the same warning and delineation of curves with design speed ranging from values equal to the approaching operating

Table 6
Value of the SPF coefficient, CMF and standard error for the variable considered.

Risk Category	λ	Error	CMF	95th upper limit
A	0.157	0.1366	1.17	1.53
B	−0.1795	0.2002	0.84	1.24
C	−0.1604	0.3372	0.85	1.65
All	**−0.0953**	**0.1065**	**0.9091**	**1.12**

speed (i.e. $\Delta V = 0$) to differences lower than 20 km/h (i.e. $\Delta V = 20$).

The estimation of CMFs for the different categories highlighted an increase in the expected average crash frequency after treatment for category A and a slight decrease for categories B and C.

Even if cross-sectional designs tend to estimate higher CMFs than those derived from before-after studies [18,29], the results clearly indicate a trend of unsatisfactory effectiveness of the SAFESTAR system application in Poland for some sites classified in category D.

The higher crash rate of category A treated curves compared to sites that have Poland's regular signs, shows that category A includes both curves which probably do not need any signs and curves where a higher warning would be proper. The uniform signing within category A may produce inconsistent driver behavior with excessive speeds on the sharper curve within the category whose design differential is higher than 45 km/h. By analogy, a revision of the signing system should be considered to exclude from category A the curve warning sign as originally proposed.

Besides, it is recommended to reconsider the comparison of design speed in curves with operating speed on tangent that should be harmonized in terms of operating speed only, and a change in the upper and lower limits of the categories which should be revised.

More specifically, boundaries should be defined to achieve a more gradual increase in CR between categories and to reduce the installation of signs where not needed.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ress.2019.03.028.

References

[1] Elvik R, Høye A, Vaa T, Sørensen M. The handbook of road safety measures. 2nd ed. Bingley, UK: Emerald Publishing Group; 2009.
[2] Lamm R, Beck A, Ruscher T, Mailaender T, Cafiso S, La Cava G. How to make two lane rural road safer - Scientific Background and guide for practical application. UK: Wit PressAshurst Lodge, Ashurst, Southampton SO40 7AA; 2007. ISBN 978-1-84564-156-6.
[3] TRL & Department for International Development. Horizontal curves, highway design note 2/01. Crowthorne, UK: TRL Ltd; 2001.
[4] Austroads & ARRB Group Ltd. Treatment type: curve widening: chevron alignment markers. Vermont South, Vic: Austroads road safety engineering toolkit, Austroads & ARRB Group Ltd; 2010.
[5] Re J, Hawkins H, Chrysler S. Assessing benefits of chevrons with full retroreflective signposts on rural horizontal curves. Transp Res Rec 2010(2149):30–6.
[6] Montella A. Safety evaluation of curve delineation improvements empirical bayes observational before-and-after study. Transp Res Rec 2009(2103):69–79. Transportation Research Board of the National Academies, Washington, D.C.
[7] Austroads & ARRB Group Ltd. Treatment type: linemarking & delineation: advisory speed signs. Vermont South, Vic: Austroads Road Safety Engineering Toolkit, Austroads & ARRB Group Ltd; 2010.
[8] Bonneson J, Pratt M, Miles J, Carlson P. Horizontal curve signing handbook, research report 0-5439-P1. Austin, Texas, USA: Texas Transportation Institute; 2007.
[9] Jurewicz C, Chau T, Mihailidis P, Bui B. From research to practice – development of rural mass curve treatment program. ARSRPE conference. 2014.
[10] SWOV Institute for Road Safety Research. Leidschendam, NL, Standard for Road Design and Redesign 2002. Final Report, November.
[11] Turner B. and Makwasha T. Austroads 2014 Publication publication Nono. AP-R449-14 methods for reducing speeds on rural roads – compendium of good practice.
[12] Montella A, Imbriani LL, Marzano V, Mauriello F. Effects on speed and safety of point-to-point speed enforcement systems: evaluation on the urban motorway A56 Tangenziale di Napoli. Accident Anal Prevent 2015;75((9) August):164–78.
[13] Herrstedt L, Griebe P. Safer signing and marking of horizontal curves on rural roads. Traffic Eng Control 2001;42(3):82–7.
[14] Cardoso J. Safety assessment for design and redesign of horizontal curves. International symposium on highway geometric design, 3rd, 2005, Chicago. Transportation Research Board (TRB); 2005. p. 20.
[15] Helman S, Kennedy J, Gallagher A. Bend treatments on the A377 between cowley and bishops Tawnton: final report, PPR494. Crowthorne, UK: TRL Ltd; 2010.
[16] Ambros J, Valentová V, Gogolín O, Andrášik R, Kubeček J, Bíl M. Improving the self-explaining performance of Czech national roads. Transp Res Rec 2017;2635(1):62–70https://doi.org/10.3141/2635-08.
[17] Kieć M, Ostrowski K. Wpływ niestandardowego oznakowania łuków poziomych na zachowania kierowców. Konferencja Naukowo–Techniczna Wpływ środków organizacji na bezpieczeństwo ruchu drogowego. Kielce 2005:99–108.
[18] Gross, F., Persaud, B., Lyon, L. A guide to developing quality crash modification factors. U.S. Department of Transportation and Federal Highway Administration (FHWA). Report No. FHWA-SA-10-032.
[19] Srinivasan R, Baek J, Carter D, Persaud B, Lyon C, Eccles K, Gross F, Lefler N. Safety evaluation of improved curve delineation. Washington, D.C.: Federal Highway Administration; 2009. Report No. FHWA-HRT-09-045.
[20] HSM. The highway safety manual 2010http://www.highwaysafetymanual.org.
[21] Cafiso S, D'Agostino C. Assessing the stochastic variability of the benefit-cost ratio in roadway safety management. Accident Anal Prevent 2016;93(August):189–97. https://doi.org/10.1016/j.aap.2016.04.027.
[22] Cafiso, S., D'Agostino, C. Reliability-based assessment of Benefits in roadway safety management. Transportation Research Record: Journal of the Transportation Research Board, 2513, 1–10. Washington, D.C., USA. DOI: 10.3141/2513-01.
[23] Montella A, Imbriani LL. Safety performance functions incorporating design consistency variables. Accident Anal Prevent 2015;74(January):133–44.
[24] Cafiso S, D'Agostino C, Persaud B. Investigating the influence on safety of retrofitting Italian motorways with barriers meeting a new EU standard. Traffic Injury Prevent 2017;18(3):324–9. https://doi.org/10.1080/15389588.2016.1203424. Published.
[25] Cafiso S, D'Agostino C, Kiec M, Bak R. Safety assessment of passing relief lanes using microsimulation-based conflicts analysis. Accident Anal Prevent (2017) 2017. https://doi.org/10.1016/j.aap.2017.07.001.
[26] Cafiso S, D'Agostino C, Bak R, Kiec M. Assessment of road safety for passing relief lanes using microsimulation and traffic conflict analysis RSS2015 Special Issue of ATS Adv. Transp. Stud., Int. J. 2016;2:55–64. (Volume 2 (2016):52-64 April 2016).
[27] Akaike Hirotugu. A new look at the statistical model identification. IEEE Trans Autom Control 1974;19(December (6)):716–23. https://doi.org/10.1109/TAC.1974.1100705.
[28] Hauer E, Bamfo J. Two tools for finding what function links the dependent variable to the explanatory variables. Proc., ICTCT Conference. 1997.
[29] Hauer, E. The art of regression modeling in road safety. Springer International Publishing. DOI 10.1007/978-3-319-12529-9.